

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



A FIRST FOR FLIGHT

Plane powered by an engine with no moving parts takes to the air

PAGES 476 & 532

TECHNOLOGY

DETAILED VIEW

The revolution in electron microscopy

PAGE 462

SECURITY

BLOCKCHAIN UNDER THREAT

How quantum computers could bankrupt Bitcoin

PAGE 465

GENETICS

MOSQUITO SEQUENCE

High-quality genome for vector of Zika and dengue

PAGES 482 & 501

 NATURE.COM

22 November 2018

Vol. 563, No. 7732

THIS WEEK

EDITORIALS

POSTDOCS Survey reveals alarming new abuse of young scientists **p.444**

WORLD VIEW Visa applications show United States is blocking talent **p.445**



CHEMISTRY Molecule that grabs sugar points to better diabetes care **p.447**

Flight test for ion drive

The first flight of a remarkable aircraft propelled by ionic wind could signal a future with cleaner and quieter aeroplanes.

In February 1904, a short news item in *Nature* marked a monumental event. It recorded the achievements of the American brothers Orville and Wilbur Wright and the contraption that they had launched from a hill in North Carolina a couple of months earlier. “They now appear to have succeeded in raising themselves from the ground by a motor-driven machine,” *Nature* stated. It was, “the first successful achievement of artificial flight”. That first trip lasted barely 12 seconds.

Nearly 115 years later, *Nature* reports on another historic brief flight, which this time lasted 8–9 seconds. On page 532, researchers at the Massachusetts Institute of Technology (MIT) in Cambridge describe an aviation breakthrough that will draw inevitable comparisons to that wobbly and fragile first journey by air. The aeroplane is powered by a battery connected to a type of engine called an ion drive that has no moving parts.

There are no passengers, either. The whole device — which has a 5-metre wingspan — weighs just 2.5 kilograms, about one-tenth of a typical commercial flight passenger’s baggage allowance. The aeroplane barely gets off the ground, cruising in tests at an altitude of 1.5 feet (0.47 metres). But anyone who watches the machine fly (see go.nature.com/2kk86jz for a *Nature* video) can surely see glimpses of a future with cleaner and quieter aircraft.

A News and Views article on page 476 delves into the technical details and the challenges that must be addressed to scale up the prototype plane. Is such a goal achievable? Conventional wisdom would say probably not. But then it also said that aircraft with ion-drive, or electroaerodynamic, engines — which create thrust by using electrical forces to accelerate ions in a fluid to form an ionic wind — would never fly at all. The thrust, after all, is produced only by the wind generated by the movement of ionized air molecules as current passes between two electrodes, one thinner than the other.

Ionic wind was first identified in the 1960s, but most scientists and aviation professionals since have insisted that the process was never going to be efficient enough to be useful, and left experiments to enthusiasts and hobbyists. Yet, not only do the MIT researchers demonstrate the first flight of an aeroplane propelled in this way, but they also show that the efficiency will increase as the velocity of the aircraft increases, because the electrodes that act as the engine create such little aerodynamic drag.

The scientists’ success will surely spur on others to re-explore a technology that was long forgotten. This will no doubt include military research, and some of the possible applications — silent drones and engines with no infrared signal that are thus impossible to detect — will rightly worry many and should be openly discussed.

This first flight will stimulate both awe and anxiety — just as the first powered flight by the Wright brothers did. Will it prove as influential? As you read this, between 6,000 and 12,000 commercial aircraft are airborne, and those are a fraction of the 100,000 or so flights scheduled each day. And every one of these aircraft is

sending greenhouse-gas emissions high into Earth’s atmosphere.

Predictions about the future of flight are dangerous because work can be overtaken by events or exposed as wishful thinking. (Just four years before the aerial carnage of the Second World War, *Nature* solemnly predicted that the risk of attack from the air was remote. And in the 1970s, it reported claims that a hydrogen-powered aircraft could take to the skies by the end of the twentieth century.)

“This first flight will stimulate both awe and anxiety.”

When the Wright brothers made their historic flight in December 1903, it didn’t receive that much attention. In part, that was because their idea was just one of several being explored to achieve flight — with others betting on the success of gliders, airships and even kites. The same is true today. Ion-drive engines are just one much-needed option to improve the efficiency and environmental impact of aircraft engines, alongside tweaks to fuel and design. Let’s hope some of them take off. ■

Brexit end game

The impact on science of Britain leaving the EU is still uncertain.

The *Daily Express* UK tabloid newspaper has been one of the loudest voices to argue that Britain should exit the European Union. But last week it admitted that the move could have some downsides. In what it labelled as a “BREXIT BOMBSHELL”, the paper reported the fears of industry scientists that: “Leaving the EU will be bad for UK science.”

No arguments here. Long before the 2016 referendum and the bitter disputes that have followed, it was clear that the United Kingdom’s membership of the political bloc benefits its research in numerous ways. (And European science benefits from Britain’s input, as well.) Leading figures and Nobel prizewinners have queued up to say the same. But although the *Daily Express* bombshell might not be news to *Nature* readers, its publication should remind researchers in Britain and elsewhere that what is obvious to them is not so to everyone. Some messages bear — indeed, demand — repetition, and the unnecessary damage that Brexit will cause to research is one of them. The need to continue to emphasize this message is especially crucial as political negotiations on how and when Britain severs its ties with the EU approach the end game.

The Brexit waters remain murky, but the superstructure of a possible ‘divorce’ deal that would set the terms of Britain’s withdrawal is just about visible through the gloom. Last week, the British government and

the EU produced a 585-page document that marks their joint attempt to set the terms. It's a draft of a deal, and the current alternative to a 'no-deal' scenario. The British public got to see it only after a fractious meeting of members of the ruling Cabinet had approved the wording — only for some Cabinet members to then promptly resign.

Anyone brave enough to read the full tome will find few references to science (and few were expected). It largely covers the thorny matters of Britain's divorce bill and a political and trading mechanism to avoid having to reinstate a hard border with Ireland. It leaves most of the key issues that are important for scientists — including immigration and access to funding — to form part of a future agreement on the EU–UK relationship. On this, the government released only a meagre outline. On the downside, this prolongs the uncertainty and unrest that is already affecting researchers. But for those determined to seek positives, it does mean that much remains in play — and that means scientists and their advocates must keep on keeping on about how Brexit is bad for them and for UK research, and how policymakers must find ways to limit the damage.

Among the continuing uncertainties, we still do not know whether UK-based scientists will be able to continue to draw grants from big-money EU research-funding programmes. Nor do we have any details on the likely shape of Britain's future immigration system, and thus how easily highly skilled EU citizens, including scientists, will be able to come to work in Britain. Freedom of movement between the EU and the United Kingdom, which has proved a boon to science in both directions, was not part of the deal, but in the days after its publication, Prime Minister Theresa May reiterated that, in the long term, EU citizens would enter on an equal footing with migrants from the rest of the world. (That EU citizens already in the United Kingdom should be able to remain was one welcome detail that the agreement did spell out and one that should

ease the anxieties of many researchers and their families.)

Needless to say, this journal argues that skilled scientists should be able to move to the United Kingdom after Brexit with few restrictions, and the evidence that this will benefit science should make it a political priority. A briefing document published alongside the agreement text does hint at provisions for some visa-free travel between Britain and EU countries. This is encouraging news for researchers who are used to travelling for collaborations and conferences.

The draft agreement text does place one field of British science and technology on firmer post-Brexit ground. It confirms that Britain will leave Euratom, the pan-EU nuclear regulator, and that responsibility for issues such as ensuring non-proliferation will pass to the control of the United Kingdom's own regulator. But the text adds nothing on issues that concern UK nuclear-fusion scientists, such as whether an independent Britain will be able to negotiate continued membership of the ITER fusion experiment in France.

Brexit is due at the end of March 2019. Before then, the agreement text must overcome a series of hurdles, not least a vote in the UK Parliament next month. The political landscape is highly volatile — Britain is already on its third Brexit minister since July, and hard-line Brexit supporters could yet trigger a leadership challenge to May, and possibly a general election. Meanwhile, there is growing support for a 'people's vote' on any agreement passed by Parliament — effectively, a second public referendum.

Much remains at stake. Scientists must continue to lobby for a Brexit settlement that protects and promotes research. There is still time to have a voice. ■

Protect postdocs

A survey of young scientists in the United States highlights the exploitation of visa holders.

Most of the research and analysis on the fate and experiences of young scientists focus on PhD students. This is probably because these students, in theory at least, have a broader spectrum of opportunity. Many postdoctoral researchers tend to have chosen a path to an academic career. What determines the outcome? And what happens to those who choose a different route? Better information and tracking would help to inform those making this decision.

Some useful — and worrying — research on this issue was published last month by two US academics in the journal *Research Policy*. The study is based on interviews with 97 postdocs from 5 major US research institutions, as well as 35 principal investigators (PIs), university administrators and industry employers (C. S. Hayter and M. A. Parker *Res. Pol.* <http://doi.org/cw62>; 2018). The interviews were conducted in 2016 and 2017. More than half of the postdocs (52.6%) worked in the life sciences.

Many of the issues these postdocs report are familiar: chiefly, how hard it is to land a tenured full-time position in academia. But the research also revealed a new — and alarming — complaint from a handful of these young scientists. Some PIs are exploiting the fact that overseas scientists rely on them for continued visas. The responses suggest that senior scientists are using this reliance to force postdocs to work longer hours and endure unacceptable conditions.

The following was said to the study's authors by a postdoc at a leading US university: "When I arrived at [the university] my PI explained to me that he approved my visa renewal ... he then told me

he was going to pay me 70 per cent of the salary he promised before I got here ... when I asked him if this is normal, he just asked me if I was serious about working [at the university]"

And this came from another: "Our PI creates this pressure cooker environment in our lab ... you see the foreign postdocs sleeping on the floor of the labs, working 100-plus hours a week ... PIs know what they are doing ... they take advantage of these guys."

Here is the view of a university administrator: "I see something bad almost every week and it seems to be getting worse ... postdocs come into my office and ask me if this or that seems wrong to me ... the visa issue is a big one because foreign postdocs are afraid to report their PIs ... these are small scientific communities and PIs will blackball their postdocs if you cross them."

The paper labels such behaviour as socially irresponsible, but that seems too mild. It is exploitation. It is unacceptable. And it must stop. These are anecdotal reports, and we have no way of knowing how large the problem is, or whether the increased political scrutiny of foreign visitors to the United States has changed the situation.

Most estimates agree that about half of the postdocs working in the United States are overseas visitors who rely on short-term visas. Institutions typically sponsor the renewals and extensions. This is largely done by individual departments and lab heads, with universities' central administrations having little formal role in the recruitment and experiences of postdocs. This puts senior scientists in a position of power. None should use this as leverage against less senior colleagues — many of whom are far from home and vulnerable. Colleagues who see such actions should report them.

Future assessments and surveys of postdocs should probe this issue further. "This was a qualitative study, so it's important to recognize that our findings are not generalizable to broader populations of postdocs," the study authors told *Nature*. Let's hope not. Everyone should agree with the postdoc who told the interviewers: "[I] realized that students can really be taken advantage of and this left a bad taste in [my] mouth with academia." ■



America, don't throw global talent away

Rhetoric and policy are keeping innovators out, warns **William Kerr** — specialist visa applications have fallen by one-fifth in the Trump presidency.

US President Donald Trump is part of a chilling parade of politicians — Turkey's Recep Tayyip Erdoğan, Hungary's Viktor Orbán, French National Front leader Marine Le Pen, Brazil's Jair Bolsonaro — who have risen to prominence in the past decade by fueling anti-immigrant sentiment. But when Trump is grandstanding about how illegal aliens “infest” our country, there's something he neglects to mention: immigrant success benefits the United States. As people in my country celebrate Thanksgiving this week, we should be grateful for what global talent has done for our economy.

Since 1900, immigrants have made up one-third of US recipients of Nobel prizes in chemistry, physics, medicine and economics. Immigrants account for more than one-quarter of the approximately 110,000 patents filed in the United States each year. There are more than 1 million foreign students in US universities, representing about 5% of enrollees and providing an estimated US\$39-billion annual stimulus to the economy.

The United States came to its leading position in science and technology in part because talented immigrants could thrive here, as I document in my recent book, *The Gift of Global Talent*. The global nature of US academia seeds connections and collaborations that make it stronger. The influx of scientists and engineers fleeing Nazi Germany (including Albert Einstein and computer scientist John von Neumann) remains the most dramatic example.

Researchers and entrepreneurs immigrate because the United States offers access to the global scientific frontier, from biotechnology to artificial intelligence. It has large, unified markets and mostly welcomes new arrivals: for example, Sergey Brin, co-founder of Google, was born in Russia; Dara Khosrowshahi, chief executive of Uber, in Iran; and Rafael Reif, president of the Massachusetts Institute of Technology, in Venezuela. The US economy improves and grows owing to ideas that immigrants develop.

But I fear the country is losing ground as the pre-eminent destination for tomorrow's science and technology leaders. Since 2016, applications for H-1B visas, which most foreign specialists need to work in the United States, have fallen by 19%. Foreign postgraduate-student applications to US business schools are down by 11% (compared with a 2% drop in domestic applicants). And numbers of international students have flattened out for the United States while continuing to increase for other countries, including Australia and Canada. Surveys of applicants and institutions suggest significant concerns about future US visa policy and openness to immigrants.

For all Trump's talk about his predecessors flinging wide US doors, the immigration process is onerous. It takes years to go from student visa to H-1B to citizenship, even for highly talented people in much-needed specialties. Until the past couple of years, talented people

felt the gain to be worth the pain. Now, uncertainty is eroding this cost-benefit calculus. People are most willing to invest when they are confident. Add volatility, and we become less likely to buy a home, launch a big project or relocate to a new country for education or work.

An unintended experiment shows this uncertainty bogeyman in action. In 2004, the annual cap on the H-1B visa supply reverted from 195,000 to 65,000 (today, the cap is 85,000). This made it harder for most foreign graduating students to find work in the United States. But legislative quirks left citizens of five countries (Australia, Canada, Chile, Mexico and Singapore) unaffected. Undergraduate enrolments from those nations did not change from 2000–01 to 2006–07; those from affected countries dropped by 14%.

Who was most likely to look elsewhere? The best students: standardized-test scores of applicants from affected countries declined by 20 points (1.5%). The H-1B visa cap was not intended to stop promising students enrolling in university and paying tuition fees, but it had that effect.

More-recent US policy changes make it harder for foreigners to launch businesses, have a working spouse or untangle visa problems. These will reduce the opportunity that migrants see in the United States.

But, in my view, what will do most to scare global talent away is hostile political rhetoric. In the days before the mid-term elections in early November, Trump deployed troops to the Mexican border to guard against Central American migrants, and claimed that he could retroactively remove the Constitution's provision of birthright citizenship.

Already, the antiquated H-1B visa system offers too few visas and allocates them poorly, awarding spots by lottery rather than need. In the first few days after the application window opened in April 2018, the government received 190,000 applications for 85,000 total slots, nowhere near enough for the numbers of international job applicants and the many tens of thousands of foreign students and postdocs already here.

Americans — including 55% of Trump voters, according to a 2017 poll (see go.nature.com/2zyaer) — broadly support expanding skilled immigration. Reasonable people can disagree about the optimal numbers and types of immigration. But at a time when the nation most needs this discussion, its dialogue is full of vitriol. Regardless of political affiliation: how many talented people outside the United States now think more highly of the country? How many will be less likely to do their research and make discoveries and inventions in the United States?

This is not just a lost opportunity: it spells disaster. ■

William Kerr co-directs the *Managing the Future of Work* project and is a professor at the Harvard Business School in Boston, Massachusetts. e-mail: wkerr@hbs.edu

WE SHOULD BE
GRATEFUL
FOR WHAT
GLOBAL
TALENT
HAS DONE FOR OUR
ECONOMY.

SEVEN DAYS

The news in brief

SPACE

Mars rover

NASA will send its next Mars rover to Jezero crater, it said on 19 November. The site is a 45-kilometre-wide crater, once filled with water, where Martian life could have thrived. There, the US\$2.4-billion rover will collect rock samples for eventual return to Earth by an as-yet-unplanned mission. Some scientists want NASA to send the 2020 rover to explore a second site called Midway, 28 kilometres from Jezero crater, where it could sample some of the most ancient rocks known on the red planet. Jezero beat several other potential destinations for the rover. They included Columbia Hills, which the Spirit rover explored between 2004 and 2011.

EVENTS

'NIPS' renamed

The board that runs a leading machine-learning conference has decided to stop using the acronym commonly used to refer to the event — NIPS — after a long-running row over whether it is offensive. The annual meeting — the full name of which is Neural Information Processing Systems — will now go by the moniker NeurIPS. The change will be in force at the next meeting, which starts on 2 December in Montreal, Canada. The move comes after months of mounting pressure about the name and the hostile environment that some women say they have experienced at the event in the past. In April, the NIPS Twitter account said that the board would consider a name change, after around 120 academics from Johns Hopkins University in Baltimore, Maryland, signed a letter highlighting “disappointing behavior” at the 2017 event. The letter said that the “acronym of the conference



CHIP CHIPMAN/BLOOMBERG VIA GETTY

Fishers sue over climate change

A coalition of crab fishers along the US Pacific coast has sued 30 fossil-fuel companies, alleging that they are responsible for the global warming that has damaged coastal ecosystems and affected their livelihoods. The lawsuit, filed on 14 November by the Pacific Coast Federation of Fishermen's Associations, says warming waters have spurred toxic algal blooms, shortened

fishing seasons and even closed entire fisheries. The fishers are seeking economic compensation from companies including ExxonMobil, BP, Shell and Chevron. Sean Comey, a senior adviser at Chevron, said that the suit is “without merit and counterproductive to real solutions to climate change”. ExxonMobil, BP and Shell did not respond to *Nature's* request for comment.

is prone to unwelcome puns”. It gave examples such as an unofficial pre-conference event named TITS.

POLICY

France's integrity

France's national research centre, the CNRS, announced plans on 13 November to create its first office of research integrity to investigate scientific misconduct and promote good research practice. The organization, Europe's largest basic-research agency, has some 33,000 staff members, more than 1,000 laboratories and a budget of around €3.3 billion (US\$3.8 billion).

The office was established by CNRS president Antoine Petit and will be headed by theoretical physicist Rémy Mosseri of the agency's Theoretical Physics of Condensed Matter Laboratory in Paris. Transparency of CNRS misconduct investigations is crucial, Petit said at a press conference. Researchers subject to misconduct investigations will be informed once any allegation is determined to be worth following up. Experts conducting investigations will be screened for potential competing interests. A national body, the French Office of Research Integrity, was created in 2017 to coordinate efforts

across France's research ecosystem, and will issue a road map of its plans next month.

Plagiarism problem

A study that looked at nearly 500 papers in 100 Africa-based biomedical journals found that 63% contained some form of plagiarism (A. Rohwer *et al. BMJ Open* 8, e024777; 2018). The study, published on 8 November, sampled the papers from the database African Journals Online. The authors, led by Anke Rohwer, an epidemiologist at Stellenbosch University in Cape Town, randomly selected papers published in 2016 from each of the

ROBIN MORE/NG/GETTY

100 journals and ran the final sample of 495 papers through plagiarism-checking software. They found extensive plagiarism — defined as when at least 4 linked sentences or more than 6 individual sentences had been copied — in 83 papers. They also found that only 26 of the 100 journals had a plagiarism policy on their website; just 16 stated that they used plagiarism software. The authors recommend raising awareness about plagiarism among academics, improving editorial practices and using anti-plagiarism software for journals.

Ban reinstated

China has temporarily reinstated a ban on the use of the body parts of rhinoceroses (pictured) and tigers in medicine, the government announced on 12 November. The nation's cabinet had said in October that it would legalize the trade of the endangered species and their by-products for medicine, lifting a 25-year-old ban on the practice. This prompted an outcry from conservationists. In a statement, State Council official Ding Xuedong said that the move had been “postponed after study”, and that the government “has not changed its stance on wildlife protection”. The illicit trade of



rhino horns and tiger bones is lucrative because they are sometimes used in traditional Chinese medicines.

RESEARCH

Parkinson's trial

Reprogrammed cells have been implanted into the brain of a patient with Parkinson's disease for the first time. Japanese neurosurgeons announced last week that they had created a therapy using induced pluripotent stem (iPS) cells; these are developed by reprogramming the cells of body tissues such as skin so that they revert to an embryonic-like state, from which they can morph into other cell types. Scientists at Kyoto University transformed iPS cells into precursors to the neurons that produce the neurotransmitter dopamine. A shortage of

these neurons in people with Parkinson's disease can lead to tremors and difficulty walking. In October, a team led by neurosurgeon Takayuki Kikuchi at Kyoto University Hospital implanted 2.4 million dopamine precursor cells into 12 sites in the brain of a patient in his 50s. The scientists say they will observe the patient for six months and, if no complications arise, will implant another 2.4 million dopamine precursor cells into his brain. The team plans to treat six more patients with Parkinson's disease to test the technique's safety and efficacy by the end of 2020. (See go.nature.com/2ttrkb2 for more.)

Crisis in psychology

A large-scale effort to replicate results in psychology research has rebuffed claims that failures to reproduce social-science findings might be down to differences in study populations. The drive recruited labs around the world to try to replicate the results of 28 classic and contemporary psychology experiments, drawing study samples from 60 different labs. Only half of the experiments were reproduced successfully using a strict threshold for significance set at $P < 0.0001$ (the P value is a test for judging the strength

of a result). Studies that did not replicate varied little across samples, debunking a common theory for failure (R. A. Klein *et al.* Preprint at PsyArXiv <http://doi.org/cw63>; 2018).

AWARDS

China prize

Fifty early- and mid-career scientists in China will be awarded generous research grants under a new privately funded programme. Tencent, the Shenzhen-based communications giant, announced on 9 November that it will invest an initial 1 billion yuan (US\$144 million) in the Xplorer Prize for researchers aged under 45. Winners will be selected annually by a committee of Chinese science leaders and will receive 600,000 yuan each year for 5 years. The 50 prizes will be awarded across 9 categories, including physics, mathematics, life sciences and energy. The award is intended to give scientists a say in choosing the most promising directions for future research. Some Chinese government awards have been criticized for political interference, and for being based on past accomplishments rather than future potential. The first winners will be announced in July 2019.

SOURCE: C. BEAUDRY ET AL., THE NEXT GENERATION OF SCIENTISTS IN AFRICA (AFRICAN MINDS, 2018).

TREND WATCH

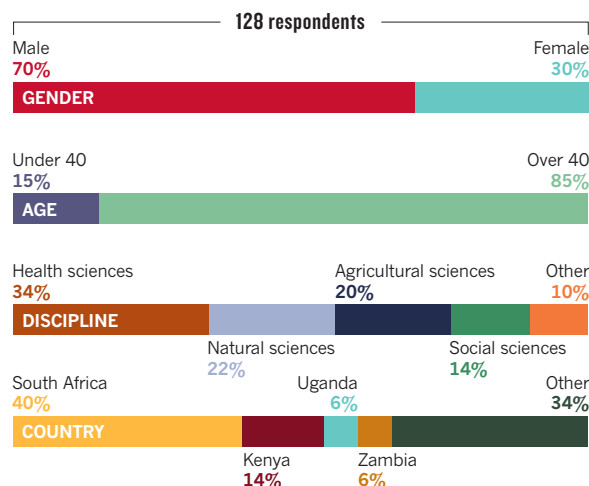
Africa's research system has created few well-funded scientists: only 2% or so of them from just a handful of countries and fields report receiving more than US\$1 million in grants over a three-year period. Meanwhile, almost half did not report receiving any research funding. Those best-funded scientists tend to work in fields and countries that rely most heavily on financing from agencies based in Europe, the United States and China, which still dominate research funding in Africa, says a report published on 6 November, called *The Next Generation of Scientists in*

Africa. The report is based on a four-year international study jointly funded by the Robert Bosch Stiftung foundation in Germany and the International Development Research Centre in Ottawa, Canada.

The authors surveyed 5,700 African researchers between May 2016 and February 2017 and analysed papers listed in the Web of Science that had authors at African institutes and were published from 2005 to 2016. Scientists' median funding over the three years before the survey was just \$5,000, although 128 researchers reported receiving more than \$1 million.

AFRICA'S TOP-FUNDED SCIENTISTS

Of 128 respondents who said they had received more than US\$1 million over the past 3 years, most were male, over 40, based in just four countries and worked in health or natural sciences.

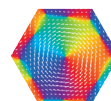


NEWS IN FOCUS

POLITICS Draft Brexit deal leaves much uncertainty for science **p.452**

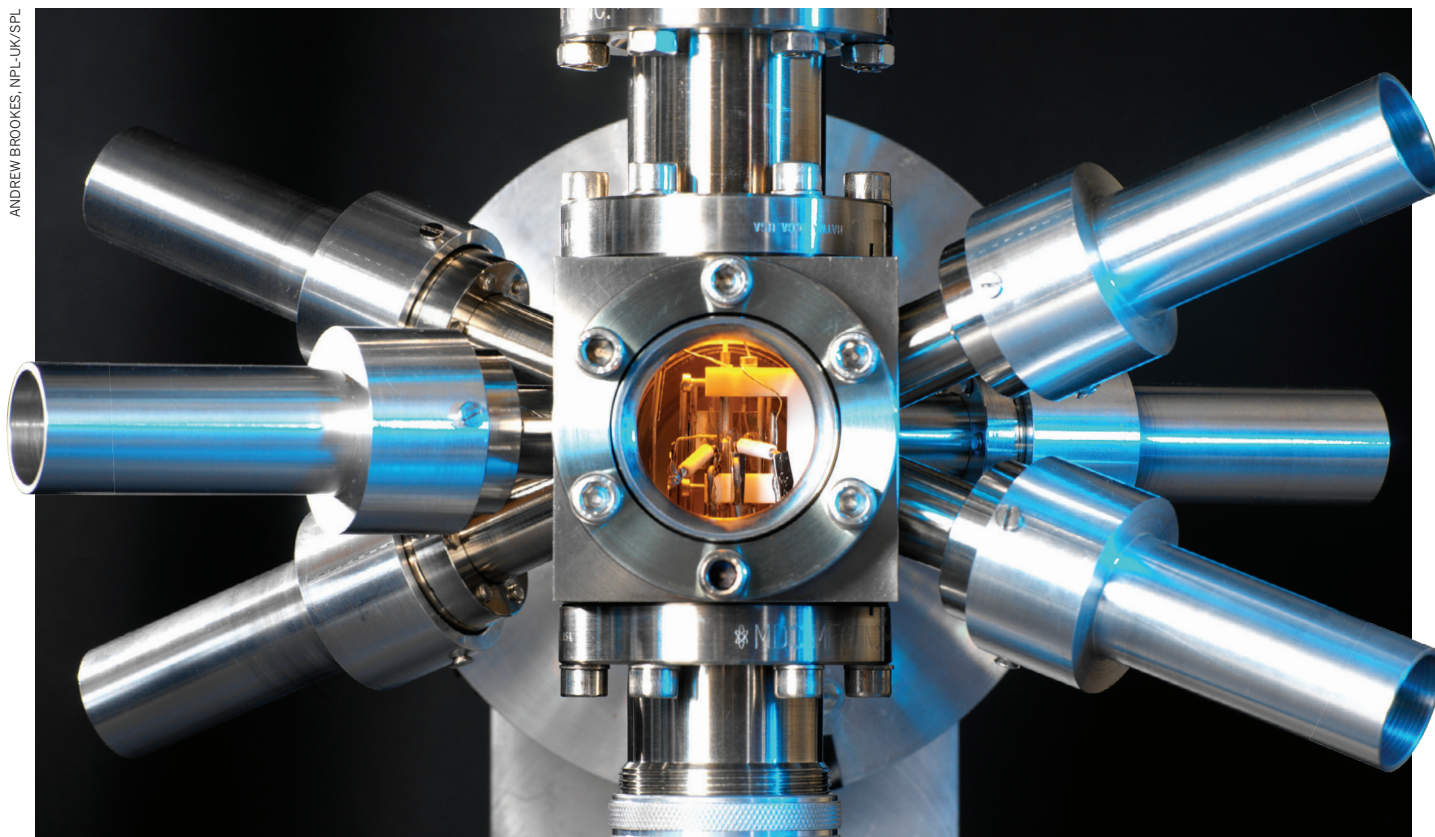
GENE EDITING UN considers banning controversial 'gene drives' **p.454**

CLIMATE Antarctic scientists begin hunt for sky scrubber **p.455**



MICROSCOPY A revolution in electron microscopes is opening new worlds **p.462**

ANDREW BROOKES, NPL-UK/SPL



The strontium clock is among the instruments that might be used to define the second in the near future.

MEASUREMENT

Metrologists ditch last physical standard units

Meet the new ampere, kilogram, kelvin and mole, now courtesy of nature's constants.

BY ELIZABETH GIBNEY

In the biggest overhaul of the international system of measurement since 1875, 60 delegates from governments around the world have voted to redefine four basic units — the ampere, the kilogram, the kelvin and the mole. The new definitions will come into force on 20 May 2019. The unanimous 16 November vote at the General Conference on Weights and Measure in Versailles, France, was met with a

standing ovation and champagne. "This is big," says Zeina Kubarych, a metrologist at the US National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland. "It's the best thrill ride you can get in metrology."

Under the new International System of Units (known simply as SI), all measures will be described using fundamental constants of nature and be derived through experiments, severing the last links between the SI and physical objects or arbitrary references.

The move allows units to be generated from their definition anywhere in the world, and to remain unassailably stable. The vote is the culmination of decades of work, and a triumph for metrologists. It means that the antique, palm-sized cylinder of metal 'Le Grand K', which resides in a vault in Sèvres on the outskirts of Paris will lose its special status as the entity that has defined the kilogram since 1889.

These changes do not mean that the SI system is perfect. Metrologists will now ►

► turn their attention to tying the definition of the second to more-precise clocks, and potentially add more units to the system. “It’s like at the end of the Harry Potter series. Good triumphs, but everything’s just trashed,” says David Newell, a physicist at NIST. In the SI, he adds, “there’s still a whole load of mess that still needs to be cleaned up”.

FOCUS ON THE CONSTANT

Measurements must always be made against a reference, and standard references ensure that units are comparable and consistent across the world — from measuring milligrams of drugs to the timing of Global Positioning Systems. The idea of basing all units on constants of nature has been around since the late nineteenth century. But it has taken almost 150 years for scientists to measure the values with enough accuracy to do so.

Metrologists working on electricity have refined experiments that count the flow of individual electrons, allowing them to use the charge on a single such particle to determine the ampere — replacing a definition that is based on a hypothetical experiment involving two infinitely long wires, which in reality can only be approximated. The kelvin will soon be defined by the Boltzmann constant, which links energy and temperature, rather than in reference to conditions at a specific temperature of water, known as the triple point.

Meanwhile, the mole — long measured as the number of atoms in 0.012 kilograms of carbon-12 — will soon equal the number of particles specified by Avogadro’s number.

In the case of the kilogram, redefinition meant measuring with exquisite precision Planck’s constant, a number that defines the size of packets of energy at the quantum scale. One method, known as a Kibble balance, derives Planck’s constant by weighing a known mass against an electromagnetic force. Another counts the atoms in two spheres of silicon-28 to derive a value for Avogadro’s number, which is converted to Planck’s constant.

Teams applying the two different methods only reached values that were accurate and in

“It’s like the end of the Harry Potter series. Good triumphs, but everything’s just trashed.”

close enough agreement in 2015. “The fact that they agree to a few parts in 10 million is absolutely extraordinary, as they are definitions based on completely different areas of physics,” says Terry Quinn, former head of the International Bureau of Weights and Measures (BIPM).

Because physical artefacts are vulnerable to being lost or damaged, the change makes the mass definition more reliable. Although Le Grand K has, by definition, always weighed exactly 1 kilogram, its mass has changed slightly relative to copies. It has been impossible to say whether Le Grand K loses or gains atoms, but future studies should now reveal that. A beauty of the system is that any experiment — once international comparisons have shown it to be accurate — can be used to determine the unit, says Estefanía de Mirandés, a metrologist at the BIPM. This not only makes

the system more democratic, it also ‘future-proofs’ the definitions, so that they can be used with more-precise experiments in the future, she says, potentially unlocking new technologies. Already, it allows measurements of very large and very small masses with much greater precision than today, she explains.

The second is currently described in relation to the frequency of microwave light absorbed and emitted by caesium-133 atoms. These atoms are now surpassed by ‘optical clocks’, which use different atoms that interact with higher-frequency visible light and seem to be able to keep time with less error: just 1 second over the age of the Universe. To update the definition of the second in 2026, as many metrologists hope will happen, the community will need to develop methods to compare optical clocks around the world and decide which atom, or atoms, to use as the standard.

Another bugbear that metrologists might try to resolve is finding a smoother way to include dimensionless quantities — such as the radian, the ratio of the length of an arc of a circle to its radius — in the SI system. “In some communities, there’s a huge push for that,” says de Mirandés.

For the BIPM, which was founded in 1875 to host the physical kilogram and metre standards, the SI revolution is bittersweet. Speakers at the meeting cheerfully quipped that there is no need to go to Paris any more. The BIPM now hopes to forge a role making comparisons between worldwide realizations of units, to ensure their accuracy, says de Mirandés. “It’s the end of a period, but also the start of a new one.” ■

POLITICS

Brexit divorce deal divides politicians and scientists

Draft treaty confirms Britain will leave European nuclear body — but uncertainties remain.

BY ELIZABETH GIBNEY & HOLLY ELSE

After two years of negotiations, the first real glimmers of what Brexit might involve have emerged. On 14 November, the Cabinet, the UK government’s senior decision-making body, backed a draft agreement that defines the terms of the country’s withdrawal from the European Union.

For science, many of the specifics that will be most relevant are still to be thrashed out. The 585-page exit treaty, if approved, largely confirms previous commitments made by the UK government, and mostly outlines what will happen during the transition period that begins

after Britain leaves the bloc on 29 March 2019 and finishes at the end of 2020.

The text offers details on the future of nuclear regulation in the United Kingdom — but it has little to say on immigration or how access to valuable EU research funds might change. These details are likely to form part of a later agreement that will define the future UK–EU relationship, and which will be negotiated only after March 2019.

An outline of the structure of this relationship, sketched out in a short accompanying document, hints at the possibility of visa-free travel for short visits to and from EU countries after Brexit — encouraging news for researchers

who are used to travelling for collaborations and conferences.

WHAT’S IN THE EXIT DEAL?

Hammered out in fraught negotiations with EU officials, the withdrawal agreement would allow EU citizens currently living in Britain, and their families, to claim permanent residence. This should ease fears expressed by many EU nationals resident in the country, including many scientists, that they would have to leave their jobs after Brexit.

The agreement also confirms that Britain will leave the European Atomic Energy Community, Euratom, when it pulls out of the EU.

NEUROSCIENCE

‘Mini brains’ show human-like activity

Electrical patterns resemble those in premature babies.

BY SARA REARDON

Mini brains’ grown in a dish have spontaneously produced human-like brain waves for the first time — and the electrical patterns look similar to those seen in premature babies.

The advance could help scientists to study early brain development, and many are excited about the promise of these ‘organoids’. But it also raises ethical concerns about creating miniature organs that could develop consciousness.

Researchers led by neuroscientist Alysson Muotri of the University of California, San Diego, coaxed human stem cells to form tissue from the cortex — a brain region that controls cognition and interprets sensory information. The group presented the work at the Society for Neuroscience meeting in San Diego this month.

Muotri and his colleagues grew hundreds of brain organoids in culture and continuously recorded electrical patterns, or electroencephalogram (EEG) activity, across the surface of the mini brains. The scientists were surprised by the EEG patterns that they observed.

In mature brains, neurons form synchronized networks that fire with predictable rhythms. But the organoids displayed EEG patterns that resembled the chaotic bursts of synchronized electrical activity seen in the brains of premature babies born at 25–39 weeks post-conception. Muotri is quick to caution, however, that the organoids aren’t close to being human brains.

The work is preliminary, says Hongjun Song, a developmental neuroscientist at the University of Pennsylvania in Philadelphia. But the similarities to preterm-infant EEG patterns suggest that the organoids could eventually be useful for studying brain-development disorders, such as epilepsy or autism, he adds.

But not everyone agrees. Just because the organoids’ brain waves look like those in premature babies doesn’t mean they’re doing the same thing, says Sampsa Vanhatalo, a neurophysiologist at the University of Helsinki.

And the ethical questions that this project raises about whether organoids could develop consciousness will be difficult to resolve, says neuroscientist Christof Koch of the Allen Institute for Brain Science in Seattle, Washington. Researchers don’t even agree on how to measure consciousness in adults, or when it appears in infants, he says. ■



UK Prime Minister Theresa May is seeking parliamentary support for the draft Brexit deal.

It fleshes out commitments made in a joint statement last December that Britain will be responsible for international nuclear safeguards in its own territory, in line with the existing regime overseen by Euratom.

But the document doesn’t address a key concern for some researchers: whether Britain can retain membership of the nuclear-fusion experiment, ITER, in France, which it currently has through Euratom.

Nor does it indicate whether the UK-based test bed for this project — the Joint European Torus near Oxford, which is largely EU-funded — will receive any cash after its current contract expires at the end of this year.

The agreement confirms that, during the transition period, UK scientists will remain eligible for grants under the Horizon 2020 research-funding programme until the programme ends.

And in statements following the deal’s announcement, UK Prime Minister Theresa May offered some information on the possible shape of Britain’s post-Brexit immigration system. She confirmed that ‘free movement’ between the United Kingdom and the bloc — something that researchers say has fuelled scientific collaboration — would end, and suggested that Britain would shift to a skills-based immigration system that would not offer priority to EU citizens over those from the rest of the world.

POLITICAL BACKLASH

For research, the deal “looks pretty good if we have to proceed with Brexit”, says Alastair Buchan, a pro-vice-chancellor of the University of Oxford.

However, Nobel-prizewinning geneticist Paul Nurse says that the agreement is disappointing for UK scientists. He welcomes the certainty it offers EU citizens living in the United Kingdom, but laments the lack of information about whether highly skilled EU scientists will be able to work in the country in the future.

The agreement has divided politicians, and turmoil in Parliament has cast doubt on whether it will pass the next step of approval

“The threat of a chaotic no-deal Brexit cannot be considered an option.”

required by the United Kingdom — a vote in Parliament, slated for December. The potential resulting ructions would increase the prospect of Britain crashing out of the EU without any kind of agreement on a future relationship — a situation widely feared by the science community.

“The threat of a chaotic no-deal Brexit cannot be considered an option,” says Venki Ramakrishnan, president of the Royal Society in London.

At an emergency summit of the European Council to be held on 25 November, the leaders of the 27 other EU countries are expected to formalize the agreement.

If May’s support holds, and the deal is approved by Parliament, it must then go before the European Parliament and garner the approval of a majority of member states.

If Parliament rejects the deal, Britain and the EU could go back to the negotiating table, but would have only until the departure date to agree on new terms. ■ [SEE EDITORIAL P.443](#)

‘Gene drive’ ban back on table — worrying scientists

United Nations body will again discuss risks of divisive technology, which could fight diseases.

BY EWEN CALLAWAY

Government representatives from nearly 170 countries are considering whether to temporarily ban the release of organisms carrying gene drives — a controversial technology that can quickly propagate a chosen gene throughout a population. The technique has the potential to eradicate disease, control pests and alter entire ecosystems, but with unpredictable consequences — leading some groups to call for a global moratorium on its field applications.

Chances are slim of a ban being approved at this month’s meeting of the United Nations Convention on Biological Diversity (CBD), which began on 17 November in Sharm El-Sheikh, Egypt. That’s because such a decision must be agreed by consensus, and biotechnology-friendly countries are unlikely to agree to such restrictions. Even so, some scientists worry that the discussions could set the tone for future limits on the use of the technology.

In an open letter on 14 November, a group of more than 100 researchers — including many studying gene drives — urged governments to

reject the moratorium, echoing a statement issued by Britain’s Royal Society this month.

“An open-ended moratorium on gene drives, without defining what is meant by ‘gene drive’ — that’s awfully crude and completely wrong-headed,” says Austin Burt, an evolutionary geneticist at Imperial College London who plans to attend the CBD meeting. He leads the Target Malaria project, which hopes to use the technology to control the spread of malaria by mosquitoes in sub-Saharan Africa. “It would stifle research,” he adds, because funders might cut support.

But Jim Thomas, co-executive director of the ETC Group, a pressure organization in Ottawa, supports a moratorium on the technology while the CBD considers potential risks and benefits. “It takes the wind out of the hype that this is somehow a ready solution,” Thomas says.

Gene drives are genetic elements that ensure their inheritance by offspring, allowing even harmful gene variants to spread rapidly through a population. They occur naturally in flies, mice and other organisms. But the advent of gene-editing tools such as CRISPR–Cas9 in the past few years has helped scientists to

develop ‘engineered gene drives’ that could be applied to any sexually reproducing organism.

Organisms containing CRISPR-engineered gene drives have not been released into the wild, but their development has stoked fears that even well-meaning applications of the technology, such as attempts to reduce populations of organisms that spread disease, could have unintended consequences for ecosystems.

Burt’s team — which has received tens of millions of pounds from the Bill & Melinda Gates Foundation in Seattle, Washington — is working on gene drives in malaria-transmitting *Anopheles* mosquitoes, with the aim of reducing populations to stop disease spread. Researchers hope to target other pests with gene drives; one international collaboration wants to use them to control invasive rodents in island ecosystems.

Individual nations can regulate the release of gene drives. But the CBD, an international treaty established in 1992 and signed by 168 countries — with the notable exception of the United States — lays out principles for conservation and the sustainable use of biodiversity, and influences national laws. It already places limits on the release of living



Anopheles mosquitoes, which transmit malaria, are the centre of many gene-drive research efforts.

SOLVIN ZANKL/NAUREPL.COM

MARIO TAMAYO/GETTY

genetically modified (GM) organisms.

In 2014, the CBD convened an expert panel of scientists and environmentalists, with the goal of determining whether synthetic biology poses challenges to the treaty. Over the past few years, gene drives have risen to the top of the group's list of issues to tackle, notes Thomas, who sits on the panel.

It's not the first time the CBD has considered a ban on gene drives. At a meeting two years ago, multiple organizations, including the ETC Group, unsuccessfully pushed for a moratorium on the technology.

LANGUAGE DEBATE

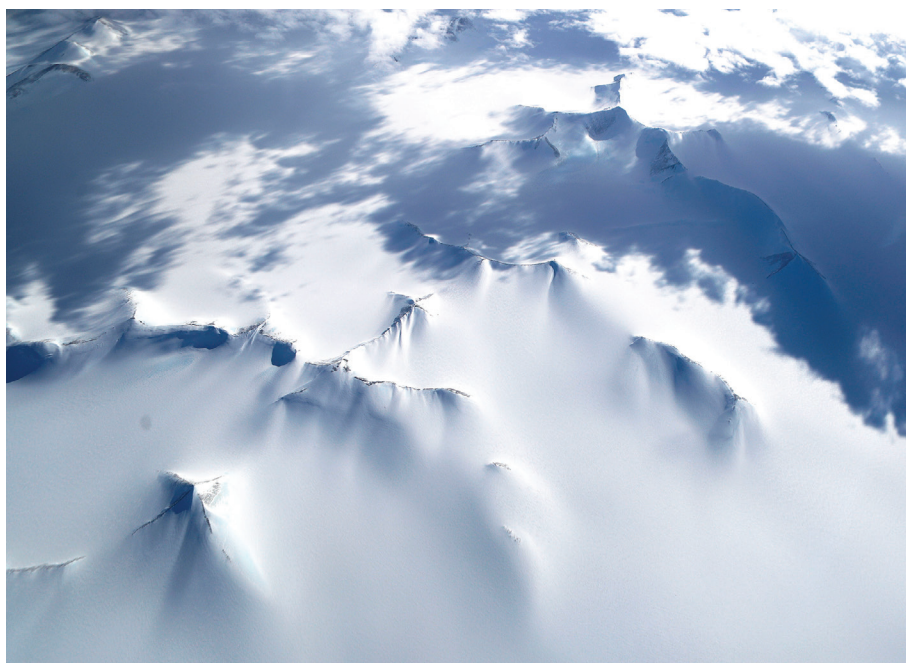
At the meeting, negotiators will again consider controversial language that calls on signatories to "refrain from the release, including experimental release, of organisms containing engineered gene drives".

Although he supports a gene-drive moratorium, Thomas expects it to face steep opposition from some countries. Canada, Australia, New Zealand and Japan have all historically lobbied against limits on biotechnologies, he notes. Any addition to the treaty must be achieved by consensus.

Even if a gene drive moratorium is not approved, the language used is likely to frame how the CBD tackles gene drives in the future. A policy document prepared by the Outreach Network for Gene Drive Research, the group that organized the scientists' letter and which includes Target Malaria, says that countries will need to decide whether to take into account positive impacts of gene drives, and how to assess the environmental risk of any releases. Target Malaria hopes to start field trials as early as 2024.

One probable outcome of the meeting is an outline for future work on policy issues raised by organisms carrying gene drives, says Todd Kuiken, a biotechnology-policy specialist at North Carolina State University in Raleigh, who is also on the CBD's synthetic-biology panel. He says that key issues include determining whether existing guidelines for assessing risks from conventional GM organisms are suitable for those carrying gene drives, and working out how to ensure that local communities potentially affected by a gene drive are consulted first.

Because it is an existing treaty signed by most countries, the CBD is likely to remain the main forum for global discussion on the topic. But Natalie Kofler, founder of a coalition called Editing Nature formed to discuss the use of gene editing in the environment, questions whether the CBD is up to the challenge. "The conversation has become very polarized, and people are seeing it as this black-or-white issue. I think it demands so much more of us," says Kofler, a molecular biologist at Yale University in New Haven, Connecticut. "I'm not sure if the CBD is providing structure to ensure a middle-ground conversation." ■



Antarctic ice traps air bubbles from Earth's pre-industrial atmosphere.

ATMOSPHERIC SCIENCE

Hunt for the sky's 'detergent' begins

Ice-core team heads to Antarctica to measure past levels of chemical that scrubs atmosphere of greenhouse gases.

BY NICKY PHILLIPS

To understand how the sky cleanses itself, a team of Australian and US researchers is heading to Antarctica to track down the atmosphere's main detergent. By drilling deep into polar ice, the scientists hope to determine how the sky's capacity to scrub away some ozone-depleting chemicals and potent greenhouse gases has changed since the Industrial Revolution — information that could help to improve global-warming projections.

The first project-members travelled to Law Dome, their drilling site in East Antarctica, this week. There, they hope to capture the first historical data on concentrations of the dominant atmospheric detergent, the hydroxyl radical. This highly reactive molecule, made of an oxygen atom bonded to a hydrogen atom, breaks down about 40 gases in the air. They include methane and hydrofluorocarbons, but not the most prevalent greenhouse gas — carbon dioxide.

Researchers have used other atmospheric gases to infer the abundance of hydroxyl over the past four decades, but chemists still refer to the radical as 'the great unknown'.

"We have been more or less in the dark when it comes to how hydroxyl has evolved from pre-industrial times to present day," says Apostolos Voulgarakis, an environmental scientist at Imperial College London. "This new research endeavour can provide unprecedented information on hydroxyl variations in the deeper past, which is exciting."

Over two and a half months, the team will drill at least two ice cores — three if time allows — to depths of about 230 metres. They will then melt the cores to extract bubbles of air that were trapped as the ice froze. The samples will represent the atmosphere back to about 1880, before emissions of greenhouse gases from human activity started to increase.

Hydroxyl radicals form naturally in the atmosphere in a reaction involving ultraviolet rays, ozone and water vapour. But because the radicals last about a second before they react with other gases and break them down, as a proxy, the team will instead measure the tiny fraction of carbon monoxide that contains the carbon-14 isotope.

Carbon-14 in carbon monoxide is produced in the atmosphere by cosmic rays at a known rate, and is almost entirely removed ►

► by hydroxyl. Because of this, scientists can use the trend in its abundance to infer the trend of the radical, says David Etheridge, an atmospheric chemist at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Aspendale, Australia, and a co-leader of the drilling project.

RISKY BUSINESS

But measuring levels of carbon monoxide that contain carbon-14 is tricky, because there are only a few kilograms of it in the atmosphere, says Etheridge. “And we’re trying to measure a bit of that over the last 150 years in the Antarctic ice.”

There is also a risk that the ice cores will become contaminated with external sources of carbon-14 from cosmic rays. This high-energy radiation cannot penetrate the ice, but the moment the cores are removed, they are at risk of exposure. This would interfere with the signal the team is trying to measure, says co-leader Vasilii Petrenko, an ice-core scientist at the University of Rochester in New York. To avoid that risk, the researchers will melt the ice and extract the air on site.

Organizing the equipment to do this and transporting it to a remote ice sheet has been a huge logistical challenge, says team member Peter Neff, an ice-core scientist at the University of Washington in Seattle.

Tractors pulled giant sleds loaded with equipment to the Law Dome drilling site, more than 130 kilometres from the nearest research station. And it will take the team 36 days to melt the ice they need to get enough air samples. “It’s a marathon, not a sprint,” says Neff.

The project is co-funded by the Australian Antarctic Division and the US National Science Foundation.

Once the researchers return from Antarctica, to assess the levels of carbon-14 in carbon monoxide, the team will convert the gas into carbon dioxide and then into graphite, from which the isotope can be measured. The scientists can then use the information to infer how hydroxyl levels in the Southern Hemisphere have changed over time.

Up to now, information on historical trends in hydroxyl levels has come solely from atmospheric models; these simulations suggest that concentrations remained fairly stable from

1850 until the 1970s, when they started to rise (A. Voulgarakis *et al. Atmos. Chem. Phys.* **13**, 2563–2587; 2013). The increase was mainly because of a boost in atmospheric warming at the time, says Voulgarakis.

The data collected from Law Dome will help to determine whether the atmospheric models have captured this trend correctly, says Matt Woodhouse, a climate modeller at CSIRO, who will use the information to improve Australia’s global chemistry-climate model, called ACCESS. “Our ability to resolve hydroxyl won’t revolutionize climate models, but it’ll increase our confidence in them.”

And accurate pictures of hydroxyl’s historical and current atmospheric concentrations are essential for developing better projections of its future levels, says Voulgarakis. This will then enable more-accurate projections of the future abundance of gases that affect climate — such as methane, ozone in the lowest layer of the atmosphere, and aerosols — that hydroxyl scrubs from the sky, he says. This would make it easier to determine the gases’ potential contribution to global warming. ■

GEOLOGY

Volcano algorithm predicts Etna’s eruptions

System tracks low-frequency waves to determine when volcano is about to erupt.

BY SHANNON HALL

Smoke filled the cabin as the Boeing 747 plunged towards snow-covered mountains in southern Alaska. All four engines had shut down, and it took the pilots eight long minutes to regain control of the aircraft. No one on board was hurt — but they had had a very close call with an erupting volcano. The jet had flown through an ash cloud.

Incidents such as this near miss from 1989 show why geologists have long sought to forecast volcanic eruptions: to protect people, whether in the air or on the ground. Now scientists are one step closer to this goal.

Maurizio Ripepe, a geophysicist at the University of Florence in Italy and his colleagues have created the world’s first automated volcano early-warning system, which alerts authorities near Mount Etna in Sicily about one hour before an eruption. The team described the system in a study published last month (M. Ripepe *et al. J. Geophys. Res. Solid Earth* <http://doi.org/cw6w>; 2018).

The approach relies on the fact that

volcanoes are noisy. Their rumblings and explosions can sound like a jet engine or even a high-pitched whistle, but they also produce low-frequency infrasound waves that people cannot hear. Unlike seismic waves, infrasound waves can travel for thousands of kilometres, allowing scientists to spot volcanic eruptions from afar. When Krakatoa, in Indonesia, erupted in 1883, its infrasound signal travelled around the globe twice.

With that in mind, Ripepe and his colleagues turned to Mount Etna, Europe’s largest active volcano. At first, they wanted to create a simple system that could detect an eruption using data from an existing array of infrasound sensors, and automatically alert authorities. But their ambitions grew when they discovered that the volcano often produces infrasound waves before it erupts, making prediction possible.

Although the finding was a surprise, the

scientists say that it makes sense, given that Mount Etna is an ‘open-vent’ volcano with exposed magma. As gas rises out of the magma before an eruption, it causes air in the volcano’s crater to slosh back and forth — creating sound waves like those in a woodwind instrument. And just as the sound of a musical instrument depends on its shape, the geometry of a crater also affects the sounds it can produce.

The team created its early-warning system in early 2010 and tracked its performance during 59 eruptions over the next 8 years. The system — an algorithm that analyses infrasound signals from the sensor array — predicted 57 of those events and sent messages to the scientists about 1 hour before an eruption took place. In 2015, the scientists programmed the system to send automatic alerts to the Italian Civil Protection Department.

An automated alert system can broadcast warnings faster than can predictions that require experts to vet information beforehand, says John Lyons, a geophysicist at the Alaska Volcano Observatory in Anchorage. And time is of the essence for communities near

“If there is an ash cloud that has suddenly popped up, then the pilots need to know that information.”



Mount Etna in Sicily is one of the world's most active volcanoes.

SANDRO SANTOLU/
FOCUS/EYEVINE

volcanoes, or passengers in a jetliner that can fly faster than 800 kilometres per hour. “You’re covering a lot of ground really fast, so if there is an ash cloud that has suddenly popped up, then the pilots need to know that information

as soon as possible,” he says. “Every minute counts.”

Although Lyons worries about the potential for false alarms, he says that the system is a pivotal step forward — not only for Etna,

but also perhaps for similar volcanoes around the globe.

These could include Kilauea, an open-vent volcano on Hawaii’s Big Island whose months-long eruption this summer destroyed whole neighbourhoods, says David Fee, a geophysicist at the Alaska Volcano Observatory. But he says that Kilauea differs from Etna in some key ways. Eruptions at Kilauea can originate from the volcano’s summit and from an area on its flank called the East Rift Zone. Etna erupts only from its summit.

Because of this, Lyons says, Mount Pavlof in Alaska, one of the most active US volcanoes, could be a better test for an early-warning system. Pavlof, whose structure resembles Etna’s, has shown increased infrasound activity before the most energetic phase of its eruptions. Its frequent activity could also give researchers a large data set with which to tune their algorithm for predicting eruptions.

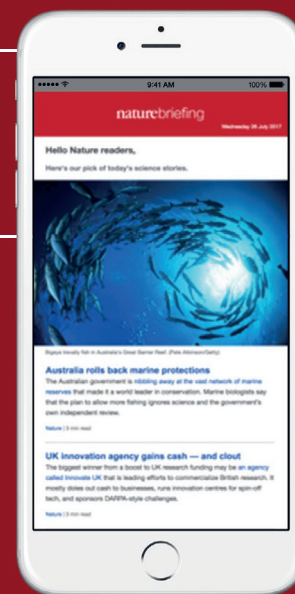
Ripepe and his colleagues are beginning to test their early-warning system in Iceland. Working with the Icelandic Meteorological Office in Reykjavik, the scientists have installed five sensor arrays across the island to monitor infrasound waves from multiple volcanoes. Among them is the infamous Eyjafjallajökull, whose last eruption, in 2010, shut down air traffic across northwestern Europe for weeks. ■

nature briefing

**What matters in science and why –
free in your inbox every weekday.**

The best from *Nature’s* journalists and other publications worldwide.
Always balanced, never oversimplified, and crafted with the scientific
community in mind.

SIGN UP NOW
go.nature.com/briefing



nature



WHY EXTREME RAINS ARE GETTING WORSE

The latest climate simulations are showing that storms will get wetter and more erratic as the world warms.

The downpour began on 13 September, when the centre of Hurricane Florence was still hundreds of kilometres from North Carolina's coast. As the giant storm lurched towards land, officials ordered more than 1.5 million people to evacuate, warning of "life-threatening" damage. On 15 September, Florence finally crashed into the United States, where it slowed to a crawl and unleashed even heavier rains. In some places, the deluge continued non-stop for four days.

By the time it was all over, Florence had dumped record amounts of rain — including nearly one metre in the town of Elizabethtown, North Carolina — and caused catastrophic flooding. Dozens of people died, and the storm racked up tens of billions of dollars in damages. Even now, months later, the area is struggling to recover.

The story of how Florence brought a thriving region to its knees is about to get a lot more familiar. Climate scientists expect that as global temperatures rise, much more rain will fall in extreme storms. The warmer the atmosphere, the more moisture it can hold, which means storms can get wetter. Even before Florence made landfall, a team based at Stony Brook University in New York predicted that the hurricane's

BY ALEXANDRA WITZE

heaviest rains would dump at least 50% more precipitation than would have happened people not warmed the planet.

Extreme rains — along with the flooding, landslides and other devastation they cause — are some of the deadliest weather events worldwide. This year, heavy rains in the Indian state of Kerala killed more than 470 people, and flooding in southwestern Japan left more than 200 dead. In the United States, flooding, severe storms and tropical cyclones account for 9 of the 11 natural disasters that have topped US\$1 billion in damages so far this year. But forecasting how the most punishing rains might change in the future has been notoriously difficult, because scientists can't easily simulate these storms in computer models.

Now, many research teams are making advances in understanding the future of extreme precipitation across the world, thanks to models with very high resolution that can provide insight into how storms evolve. Some of the most sophisticated forecasts suggest that as the globe warms, more rains will fall in severe, intermittent storms rather than in the kind of gentle soaking showers that can sustain crops. Other research indicates that the ways in which thunderstorms organize

JONATHAN BACHMAN/REUTERS

Hurricane Harvey in 2017 set records for rainfall.

themselves could change fundamentally, leading to bigger and more-powerful storms that could mean more flooding. All that makes Florence, the Indian disaster and other devastating downpours a probable glimpse of the future if greenhouse-gas emissions continue to rise. “The next 20 years will be worse than the last 20 years — all indications point to that,” says Angeline Pendergrass, an atmospheric scientist at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. “And things will be completely nuts by the end of the century if we keep doing what we’re doing now.”

MORE MOISTURE

In Kerala, it all started with a wet June and July, and accelerated over the following month. The first 20 days of August brought 164% more rain than usual. Throughout the state, landslides swept into towns and buildings as sodden ground collapsed. More than 1 million people fled their homes.

The weather might not have been the only problem in this case. Some have criticized local officials for not better managing the build-up of water behind local dams. But the flooding was ultimately traced back to heavier-than-usual storms during the summer monsoon.

And it’s clear that such storms are carrying more moisture than they used to. The moisture in the air changes depending on temperature: heat that air by 1 °C, and it can hold approximately 7% more water. The Intergovernmental Panel on Climate Change has concluded that many parts of the world are already seeing increases in heavy precipitation, thanks to human-induced climate change.

“It’s just basic physics,” says Kenneth Kunkel, an atmospheric scientist at North Carolina State University in Asheville. “Big storms with large amounts of rainfall are limited by the amount of water vapour in the atmosphere. As we increase water vapour in the atmosphere, we can increase the amount of rainfall in these extreme precipitation events.”

But the precipitation story turns out to be more complicated than that. A thunderstorm is essentially a tower of upward-moving winds that feed themselves by sucking in warm air from nearby. When the air rises high enough, it cools and condenses into rain. Storms can generate their own weather, such as creating cold pools of air near the ground that trigger more convection. And climate change can amplify these effects, causing updrafts to grow stronger and wider, which pulls in more warm air from surrounding regions and leads to more rain (see ‘Heavy rain’).

This apparently happened in Hurricane Harvey — the rainiest storm in US history, which drowned much of Houston and south Texas in a \$125-billion disaster in August last year. Three separate studies have concluded that Harvey’s heaviest rains could not be explained simply by the increase of water vapour in the atmosphere^{1–3}. Climate change made it even wetter than that.

At NCAR, Pendergrass is working with global climate models to pin down what this might mean for extreme events in the future — and especially, where they might occur. She analyses how climate change is altering how heat and energy flow in the atmosphere, which changes how precipitation is spread around the globe.

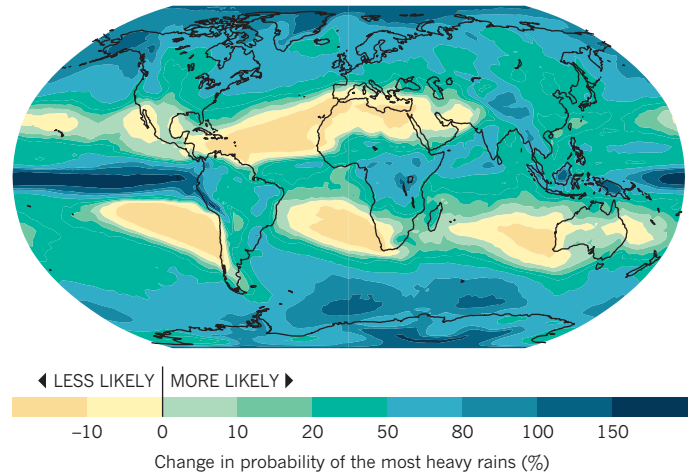
Last year, she and her colleagues reported on three computer simulations showing that precipitation is likely to become more variable across almost all land areas if temperatures rise through the rest of the century⁴. In other words, weather will get crazier: wet periods will give way to dry periods more erratically, and vice versa, across nearly all the continents.

Now, she has drilled down to study the unevenness of precipitation — that is, the difference between a light drizzle and a torrential downpour. She and Reto Knutti, an atmospheric scientist at the Swiss Federal Institute of Technology in Zurich (ETH Zurich) in Switzerland, analysed global rainfall records between 1999 and 2014. For the median of all locations in the study, it took only 12 days for half of the year’s rain to fall⁵. “Things that are considered extreme contribute a lot to the total precipitation, more than a lot of my colleagues may realize,” she says.

Looking ahead with a climate model, the researchers found that this

HEAVY RAIN

A climate model simulating daily precipitation changes suggests that if the planet warms by 3 °C, most land areas would see substantially more heavy rains.



kind of unevenness will increase. If greenhouse-gas emissions continue to climb quickly in the future, half of the extra rainfall will happen during the wettest six days of the year.

That means more deluges — and the dangers that follow. Pendergrass has been talking with water managers in Denver, Colorado, who want to know how much flooding they need to prepare their dams to handle in the future. She says other areas should prepare, too. “Rather than assuming more rain in general, society needs to take measures to deal with little change most of the time, and a handful of events with much more rain,” she and Knutti wrote last month in *Geophysical Research Letters*⁵.

Another set of simulations underscores how society needs to prepare for these swings between wet and dry. This work, led by atmospheric scientist Zachary Zobel while he was at the University of Illinois in Urbana–Champaign, took a global climate model that usually calculates conditions every 100 kilometres and forced it to a much higher resolution, of just 12 kilometres. “You need as high a spatial resolution as you can get,” says Zobel, who is now at the Woods Hole Research Center in Falmouth, Massachusetts. At a resolution of 12 kilometres, the model can reveal small-scale phenomena in the atmosphere that are important for simulating storms. But Zobel’s model requires a lot of computing power, so the team looked only at the continental United States, not the entire globe. Even using a supercomputer, it took the better part of a year for the model to crunch through its calculations⁶. But in the end, the researchers got a detailed look at how different parts of the country would be affected, in both temperature and precipitation extremes, if greenhouse-gas emissions continue to remain high out until 2094.

BOOM OR BUST

The scientists found that extreme precipitation events would increase over most of the country. They also spotted some larger changes in future patterns, such as in the position of the west-to-east jet stream that controls weather over much of the middle part of the United States.

That change arises because the Arctic is currently warming faster than are the mid-latitudes, so there is less of a temperature difference between the two regions. In response, the jet stream shifts northward in the simulations, bringing warm moist air from the Gulf of Mexico behind it. The result is that the midwestern states, where the bulk of the country’s corn and wheat crops are grown, will probably see more severe storms each spring during the planting season. Meanwhile, dry spells will grow longer, the model suggests.

“It really comes down to a boom-or-bust-type precipitation pattern,” Zobel says. “That will complicate how farmers deal with planting their crops.” The work appeared last month in *Earth’s Future*⁶.



Flooding in the Indian state of Kerala killed hundreds of people this year.

AFP/GETTY

For Andreas Prein, the future hit home in 2002, when heavy summer rains drenched central Europe. Prein was serving in the Austrian military, and his unit was dispatched to help flood-ravaged areas of northern Austria, where tiny brooks became torrents and damages reached €3 billion (US\$3.4 billion). “It was really shocking to see how little streams could do such devastation,” says Prein. “It was almost unbelievable.”

Today, Prein is an atmospheric scientist at NCAR and a leader in a new type of high-resolution climate modelling. This research aims to simulate future climate in even higher resolution than Zobel’s. Its computations narrow down to scales of four kilometres or less — which is so computationally expensive that it can be done only for relatively small regions.

Four kilometres is crucial because it’s the dimension at which individual storms evolve and grow stronger through convection. The field is known as convection-permitting climate modelling, and it allows researchers to simulate storms much more realistically. The calculations are similar to what weather forecasters do to predict how storms will develop over the next day or two. “But we want to simulate decades to centuries,” Prein says. “It’s basically copying what weather forecasting does, but on much longer time scales.”

In ongoing work, UK-based researchers have been running European-wide climate simulations at 2.2-kilometre resolution and have spotted some warning signs about future storms. “At the moment across most of Europe, the season we get extremes is primarily summer,” says team member Elizabeth Kendon, a climate scientist at the Met Office in Exeter, UK. In the simulations of a warmer world, summer downpours get heavier, and extreme events occur later in the year. That suggests officials might need to do more to prepare for flooding from winter storms.

Hurricanes are another major concern. Christina Patricola and Michael Wehner, at the Lawrence Berkeley National Laboratory in California, have used convection-permitting models to study the devastating hurricanes of Katrina in 2005 and Irma and Maria in 2017. They found that climate change had boosted rains within these storms by up to 9% — and that future warming would almost certainly mean more extreme rainfall from similar hurricanes⁷.

One of the biggest convection-permitting simulations so far, run by NCAR, covered the continental United States at a resolution of four kilometres⁸. It was actually a pair of simulations, one looking back and another peering forward. The first one simulated global climate between October 2000 and September 2013 to test how accurately the model could reproduce what happened during those years. “We started as you

would start a weather forecast, and then we just didn’t stop for 13 years,” says Prein.

The second simulation looked at a similar period at the end of the century, incorporating factors that would be expected if society continues to produce greenhouse gases at a high rate.

By comparing the two, the scientists could tease out the probable effects of increasing greenhouse-gas levels. And because the simulations were at such high resolution, they could capture how individual storms are likely to form and evolve in the future.

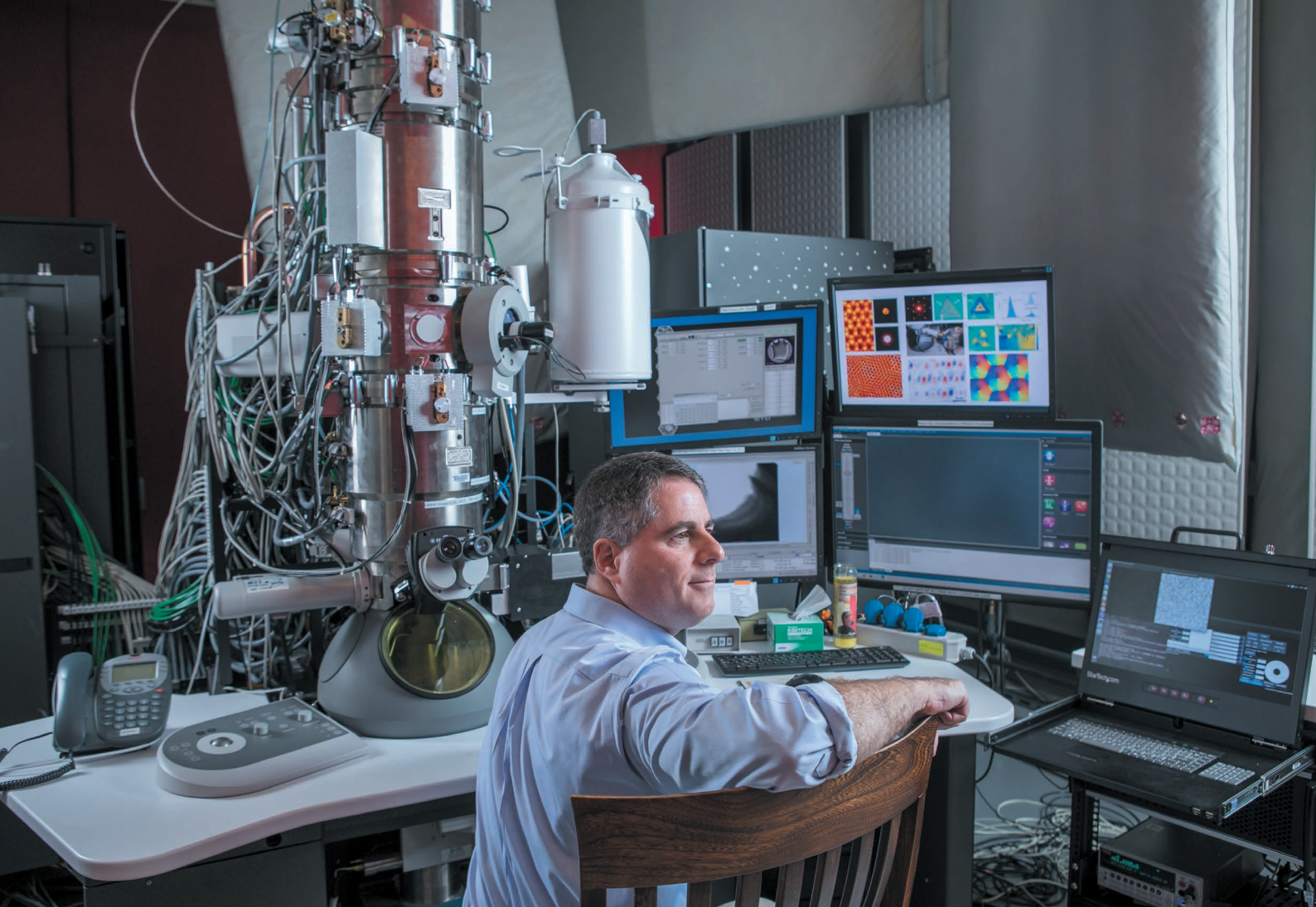
Among the many findings of such models are that intense US thunderstorms will more than triple in frequency by the end of the century, and their maximum rainfall will increase by 15–40% (ref. 9). The storms will also grow larger, almost doubling the area that would get hit with heavy rainfall. That has big implications for flood risks. “If you think about a large thunderstorm moving over a city area, it makes a big difference if the storm covers half the city catchment or all of it,” says Prein.

He and his colleagues at NCAR are now developing a second set of simulations which they hope to begin running in the coming months. It will extend the simulations northwards into Canada, aiming to study whether the powerful thunderstorms that regularly rip through the upper midwest of the United States will move into Canada in the future. The work will also cover periods lasting 20 years rather than 13, which the scientists hope will enable them to capture longer trends in changing weather patterns.

Whatever is on the way, it will almost certainly be hard to miss. Sitting in her office at NCAR on a sun-drenched morning, Pendergrass pulls up a figure showing computer simulations of what kinds of extreme precipitation might lurk in the future. “What we have seen so far is incredibly small compared to what’s coming,” she says. “And that’s kind of terrifying.” ■

Alexandra Witze writes for Nature from Boulder, Colorado.

1. Risser, M. D. & Wehner, M. F. *Geophys. Res. Lett.* **44**, 12457–12464 (2017).
2. Van Oldenborgh, G. J. *et al. Environ. Res. Lett.* **12**, 124009 (2017).
3. Wang, S.-Y. S. *et al. Environ. Res. Lett.* **13**, 054014 (2018).
4. Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C. & Sanderson, B. M. *et al. Sci. Rep.* **7**, 17966 (2017).
5. Pendergrass, A. G. & Knutti, R. *Geophys. Res. Lett.* <https://doi.org/10.1029/2018GL080298> (2018).
6. Zobel, Z., Wang, J., Wuebbles, D. J. & Kotamarthi, V. R. *Earth’s Future* **6**, 1471–1490 (2018).
7. Patricola, C. M. & Wehner, M. F. *Nature* **563**, 339–346 (2018).
8. Liu, C. *et al. Clim. Dyn.* **49**, 71–95 (2017).
9. Prein, A. F. *et al. Nature Clim. Change* **7**, 880–884 (2017).



PUSHING THE LIMITS

*Technological advances are triggering
a revolution in electron microscopy.*

Scientists can't study what they can't measure — as David Muller knows only too well. An applied physicist, Muller has been grappling for years with the limitations of the best imaging tools available as he seeks to probe materials at the atomic scale.

One particularly vexing quarry has been ultra-thin layers of the material molybdenum disulfide, which show promise for building thin, flexible electronics. Muller and his colleagues at Cornell University in Ithaca, New York, have spent years peering at MoS₂ samples under an electron microscope to discern their atomic structures. The problem was seeing the sulfur atoms clearly, Muller says. Raising the energy of the electron beam would sharpen the image, but knock atoms out of the MoS₂ sheet in the process. Anyone hoping to say something definitive about defects in the

BY RACHEL COURTLAND

structure would have to guess. “It would take a lot of courage, and maybe half the time, you’d be right,” he says.

This July, Muller’s team reported a breakthrough. Using an ultra-sensitive detector that the researchers had created and a special method for reconstructing the data, they resolved features in MoS₂ down to 0.39 angstroms¹, two and a half times better than a conventional electron microscope would achieve. (1 Å is one-tenth of a nanometre, and a common measure of atomic bond lengths.) At once, formerly fuzzy sulfur atoms now showed up clearly — and so did ‘holes’ where they were absent. Ordinary electron microscopy is “like flying propeller planes”, Muller says. “Now we have a jet.”

Muller’s images represent the latest of a burst of technological advances that are triggering a revolution in what researchers can probe using transmission electron microscopes (TEMs) — devices as tall as a room that send beams of electrons through samples to explore structures down to a size scale smaller than an atom. The machines promise to give scientists the ability to see details previously out of reach, from the structure of fragile next-generation electronics materials, to the innards of porous substances that can separate gases.

The excitement isn’t just about high-resolution images. The new capabilities also let researchers explore invisible properties of materials as never before, including electric and magnetic fields as well as hard-to-detect vibrations inside crystals. And some researchers are converting the vacuum-filled

JESSE WINTER FOR NATURE

David Muller with his team's electron microscope.

interiors of electron microscopes into tiny laboratories, so that they can study how samples behave when they are exposed to liquids and gases or varying temperatures.

A large contributor to the improvements has been speedy detectors that are sensitive to electrons. Early incarnations of these detectors have already made an impact on biology, revealing details about the construction of proteins and other substances that would be time-consuming — if not impossible — to measure through conventional X-ray crystallography. But researchers say that many of the rewards of these fresh capabilities are only now within reach — particularly when it comes to the study of nanomaterials and other synthetic systems. For a long time, people were “figuring out what you can do at all”, says Haimei Zheng, a materials scientist at Lawrence Berkeley National Laboratory in California. “I think that this field is now getting ready to address more significant questions.”

NEW RESOLUTIONS

In some ways, the electron microscope hasn't changed much since its introduction in the 1930s. The modern TEM still shoots a beam of electrons through a sample. At the far end, a detector then registers the resulting image, or researchers can use information from scattered electrons to reconstruct the sample's structure. Because electrons can have wavelengths that are thousands of times shorter than those of visible light, they are able to resolve much finer details than can an ordinary optical microscope.

Although this basic design has stayed intact, the resolving power of TEMs has improved by a factor of more than 1,000. The last big leap got its start around 20 years ago, with the emergence of electromagnets that could correct for distortions in the electron beam. By the late 2000s, these long-awaited aberration correctors had enabled advanced TEMs to reach sub-angstrom resolution.

“For materials folks, aberration correctors were a big revolution,” Muller says. “It not only let you see every kind of atom that you wanted to see, but it also let you work much quicker than you worked before.” But to take full advantage of this jump in resolution, microscopists still had to deliver intense doses of electron beams to their samples — which meant that fragile materials, including anything biological, would be damaged.

Biologists were quick to leap on another innovation. For many years, the best electronic method for taking TEM images began with radiation-sensitive scintillators, which were used to convert incoming electrons into photons that could then be detected. But the process was indirect and inefficient and led to a lot of blurring.

That changed in the early 2010s, when ‘direct-electron detectors’ became widely available. Such devices could directly and efficiently

register electrons, generating cleaner images from fewer incoming particles.

Biologists paired these detectors with frozen samples to create a TEM technique called cryo-electron microscopy (cryo-EM), which has illuminated the structures of a wide range of biomolecules. Last year, three pioneers of the approach won the Nobel Prize in Chemistry for their work.

For many materials scientists, Muller says, these detectors held less appeal. For one thing,

they couldn't tolerate many electrons per pixel, which prevented researchers from using the kind of high-intensity beam they would need to observe objects at the tiniest scales. The devices were especially ill-suited for scanning transmission electron microscopy (STEM), in which electrons are focused into a smaller, brighter beam that can then be moved across a sample. The problem was that the cryo-EM detectors were not designed to capture both the flood of electrons that pass undiverted through the sample and the small fraction that get deflected from their original path, which is crucial in STEM.

A decade ago, Muller and his colleagues began working on a detector that could nab all those electrons. Unlike those used for cryo-EM, which can have millions of pixels, the team's eventual device, called the electron microscope pixel-array detector (EMPAD), boasts fewer than 20,000 pixels. But the EMPAD is built on a half-millimetre-thick slab of silicon, so it can capture all the energy of electrons that hit it and thereby discern individual particles as well as the main beam. Muller likens the detector's million-to-one dynamic range to a back-lit picture on a sunny day. “This is a detector that would be able to get an image of all the sunspots on the Sun and the image of my friend's face in the shadow at the same time,” he says.

It was this advance that allowed Muller's team

“WE ARE MAKING STEPS IN UNDERSTANDING THE FUNDAMENTALS OF SCIENCE.”

to clearly image the MoS₂ slivers this year, with the aid of a computational method to process multiple scattering patterns, called ptychography¹. But the ability to capture all the electrons scattered by a sample gives researchers much more information to work with. Electric and magnetic fields, for example, alter how electrons are scattered. In 2016, Muller and his colleagues² showed that they could use data collected by the EMPAD to map out the magnetic field at various points in the sample — a feat difficult

to accomplish through other methods. One subject that Muller is excited to study now is skyrmions — nanometre-scale swirls of magnetism that could potentially be used to store data.

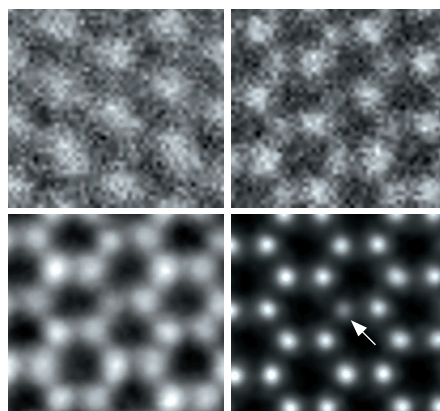
Muller's team is not the only one to create detectors with a large dynamic range. Quantum Detectors in Oxford,

UK, is one of three companies that are building electron-microscopy detectors based on Medipix, a class of chip developed at CERN, Europe's biggest particle-physics laboratory, near Geneva, Switzerland. “I think they've taken the big manufacturers by surprise,” says Damien McGrouther, a microscopist at the University of Glasgow, UK, which is working with the company. Muller, meanwhile, has licensed his technology to Thermo Fisher Scientific — a large research-supplies company headquartered in Waltham, Massachusetts.

DELICATE IMAGING

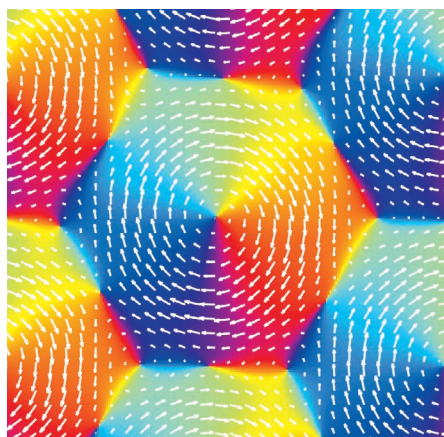
Direct-electron detectors also allow the number of electrons in a beam to be reduced — and therefore used to illuminate a range of radiation-sensitive materials. These include, for instance, metal-organic frameworks (MOFs), porous crystalline materials that researchers are exploring for many uses, such as extracting moisture from desert air and separating natural gas from other hydrocarbons. These targets can be even more sensitive to electron dose than proteins are, says Ming Pan, a physicist who works in business development at Gatan, an electron-microscopy company in Pleasanton, California. In 2017, he was part of a team that imaged a MOF at atomic resolution using one of Gatan's detectors on a TEM³.

The sensitivity and speed of direct-electron detectors, which can be faster than 1,000 frames per second, has also captured the attention of researchers working on moving electron microscopy beyond static structures. Thanks to microfabrication techniques, it is now possible to make sample holders that can do more than simply sit inside the high vacuum of an electron microscope. Researchers can control temperature, apply tension and compression, expose samples to gases and even confine liquid solutions to see how materials change in phase, structure or chemistry.

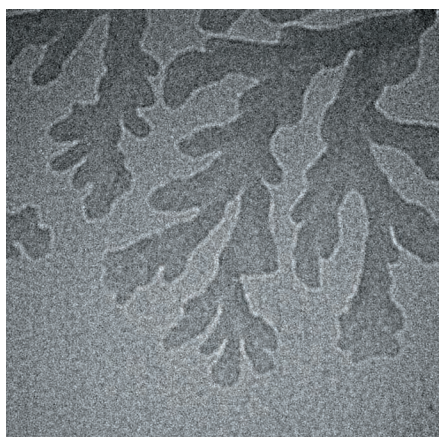


The 0.39-Å MoS₂ image (bottom right) shows a sulfur vacancy unclear in lower-resolution images.

REF. 1



Swirls of magnetism imaged in an iron–germanium film, where colour and arrows show field direction (left). Seaweed-like iron oxide nanodendrites grow on the membrane of a liquid cell in a TEM (right).



Many of these ideas aren't new, says Frances Ross, a materials scientist at the Massachusetts Institute of Technology in Cambridge. Combining through old papers, she was inspired to find discussions from the 1940s about how to look at water between two thin windows. "The ideas were out there," she says. "But they didn't have the materials, the fabrication techniques, to make it happen in a practical way."

Ross is widely credited with moving liquid cells into the practical realm. As a researcher at IBM in the early 2000s, she and her colleagues created a holder with a silicon nitride pane that was thin enough to allow electrons to pass through relatively unimpeded⁴. Since then, researchers have explored other materials for use in liquid cells, such as graphene⁵.

At the Lawrence Berkeley laboratory, Zheng is leading a multimillion-dollar US Department of Energy programme dedicated to developing the technique further. She and others have trained a variation of a detector designed for cryo-EM on liquid samples. Among other targets, they are interested in the interface between battery electrodes and electrolytes — a crucial area in which problems such as the formation of metallic filaments called dendrites can shorten a battery's lifetime, and even cause it to explode. Such studies, she says, could help in devising ways to improve performance and investigate new battery compositions. When researchers want to test materials, they often construct small batteries called coin cells to see how the ensemble performs. But, Zheng says, that cell is "almost like a black box. They don't know what is going on inside." With liquid cells, she says, researchers have a window on to the sorts of nanoscale behaviour that ultimately determine the performance of batteries, including how the dendrites grow.

Others have trained the electron microscope on more-fundamental systems. At Eindhoven University of Technology in the Netherlands, Nico Sommerdijk and his colleagues have explored the formation of fluid-filled structures that resemble the vesicles in cells. In work yet to be published, the researchers

have imaged a two-sided polymer as it self-assembles in liquid to form an artificial vesicle. And with a team led by Jim de Yoreo at the Pacific Northwest National Laboratory in Richland, Washington, Sommerdijk has studied how a polymer can bind to calcium, a process that could provide insight into how marine creatures grow the iridescent material known as nacre or mother-of-pearl. "It's not the invention of penicillin," Sommerdijk says, "but we are making steps in the fundamentals of understanding science."

Liquid-cell research has challenges. One of the biggest, says de Yoreo, is that electrons can wreak havoc when they hit water or an organic solvent, creating charged radicals that can destroy samples, shift pH or generate reducing agents that cause unintended reactions. It is also difficult to measure quantities such as pH and temperature inside the microscope.

But others are heartened by the latest research on the effect of electron beams. Patricia Abellan, a materials scientist at SuperSTEM, a research centre and user facility for advanced microscopy in Daresbury, UK, says she has seen "a revolution in the understanding of the interaction of the electron beam with matter", particularly in liquid systems. The change has been spurred in large part by collaborations with researchers who focus on studying materials affected by nuclear radiation. In the past few years, Abellan and others have explored how additives can control the growth of particles and alter pH, and how solvents other than water, such as toluene, might limit the effect of electron beams on samples in liquid⁶.

BETTER BEAMS

Advances in electron microscopy have also come from improving the electron beams themselves. Devices called monochromators have allowed researchers to narrow the range of energies for electrons that reach the sample. Researchers are starting to use that tighter spread of energy, along with spectrometers and other instruments, to reach beyond the basic structure and composition of materials and map more-sophisticated properties

at ever-finer resolutions. One such target is phonons — vibrations in the atomic lattice of materials. Mapping these vibrations at atomic resolution "would provide a lot of information on key processes behind most modern technology", Abellan says, such as how materials conduct electricity and heat.

Some researchers are turning the electron beam's potential to interfere with materials into a tool in its own right. Earlier this year, physicist Toma Susi at the University of Vienna and his colleagues used a STEM electron beam to move a silicon atom from site to site inside a hexagonal graphene lattice⁷. A similar sort of manipulation has been done for years on materials with weaker bonds in atomic-force and scanning tunnelling microscopes, Susi says, but in these cases, the results aren't stable. If the atoms aren't kept very cold, thermal energy erases the new structures. Electron microscopes are capable of higher-energy work. "Once something is manipulated," he says, "it really stays." Researchers hope that this ability may be useful for pushing atoms around inside 3D structures to, for example, create small devices for quantum computing⁸.

At the University of Antwerp in Belgium, Johan Verbeeck is looking to make electrons into a more-sophisticated probe, by passing them through plates that can alter their phase. By embedding extra information in an electron before it passes through a sample, researchers might be able to find out more about the sample's properties. "The quest is to get more information from the same electron," says Verbeeck.

Sommerdijk points to work by Nigel Browning at the University of Liverpool, UK, who has been exploring how to control a STEM beam to minimize damage. Instead of doing a comprehensive scan, a microscope could hit a subset of points in the sample. Done right, such sparse sampling could still generate a large amount of useful data. "I think it's beautiful," says Sommerdijk, adding that it could be particularly useful in liquid studies.

Muller has his eyes on other ideas; he'd like to see, for example, whether detailed materials studies can be extended from room temperature down to cryogenic temperatures — a prospect that needs more mechanical stability than electron microscopes are currently capable of. But the field is moving fast, he says. "I don't think anyone is standing still. Everyone's thinking about what do you want to build next." ■

Rachel Courtland is a features editor at *Nature*.

1. Jiang, Y. *et al. Nature* **559**, 343–349 (2018).
2. Tate, M. W. *et al. Microsc. Microanal.* **22**, 237–249 (2016).
3. Zhu, Y. *et al. Nature Mater.* **16**, 532–536 (2017).
4. Williamson, M. J., Tromp, R. M., Vereecken, P. M., Hull, R. & Ross, F. M. *Nature Mater.* **2**, 532–536 (2003).
5. Yuk, J. M. *et al. Science* **336**, 61–64 (2012).
6. Abellan, P. *et al. Langmuir* **32**, 1468–1477 (2016).
7. Tripathi, M. *et al. Nano Lett.* **18**, 5319–5323 (2018).
8. Hudak, B. M. *et al. ACS Nano* **12**, 5873–5879 (2018).

COMMENT

CONSERVATION The people and places that invented the word 'environment' **p.468**

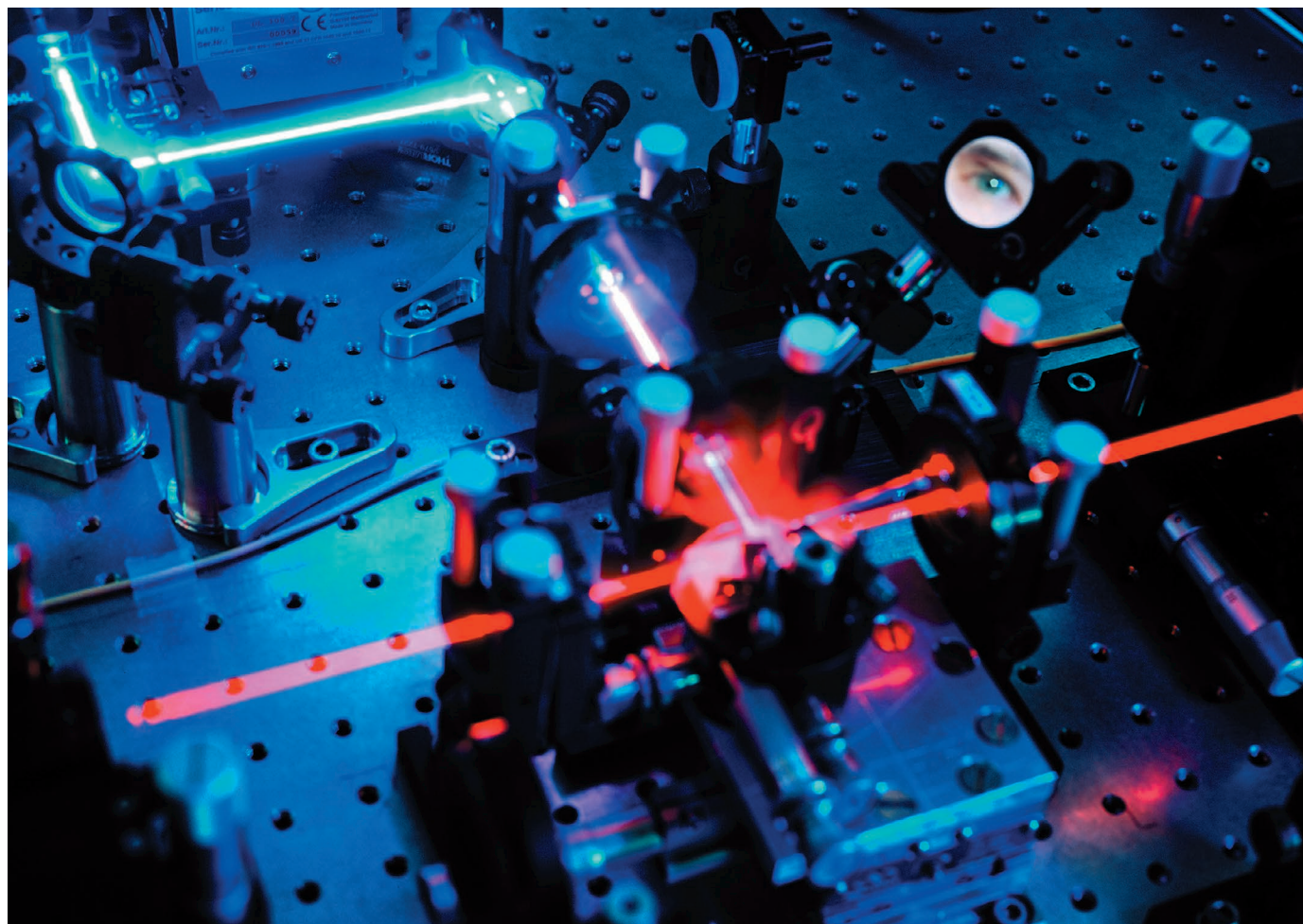
SPACE Rock legend Brian May retells the race to the Moon — in 3D **p.469**

MUSIC Celebrating the female pioneers of electronica **p.470**



OBITUARY How Paul Allen, Microsoft philanthropist, rebooted brain research **p.474**

VOLKER STEGER/SPL



Quantum cryptography equipment, which uses the principle of entanglement to encode data that only the sender and receiver can access.

Quantum computers put blockchain security at risk

Bitcoin and other cryptocurrencies will founder unless they integrate quantum technologies, warn **Aleksey K. Fedorov, Evgeniy O. Kiktenko and Alexander I. Lvovsky.**

By 2025, up to 10% of global gross domestic product is likely to be stored on blockchains¹. A blockchain is a digital tool that uses cryptography techniques to protect information from unauthorized changes. It lies at the root of the

Bitcoin cryptocurrency². Blockchain-related products are used everywhere from finance and manufacturing to health care, in a market worth more than US\$150 billion.

When information is money, data security, transparency and accountability are crucial.

A blockchain is a secure digital record, or ledger. It is maintained collectively by users around the globe, rather than by one central administration. Decisions such as whether to add an entry (or block) to the ledger are based on consensus — so personal trust ►



Conventional computer equipment inside a Bitcoin mine near Sichuan, China.

► doesn't come into it. Any party inside or outside the network can check the integrity of the ledger by making a simple calculation.

But within a decade, quantum computers will be able to break a blockchain's cryptographic codes. Here we highlight how quantum technology makes blockchains vulnerable — and how it could render them more secure.

ONE-WAY CODES

Blockchain security relies on 'one-way' mathematical functions. These are straightforward to run on a conventional computer and difficult to calculate in reverse. For example, multiplying two large prime numbers is easy, but finding the prime factors of a given product is hard — it can take a conventional computer many years to solve.

Such functions are used to generate digital signatures that blockchain users cite to authenticate themselves to others. These are easy to check and extremely difficult to forge. One-way functions are also used to validate the history of transactions in the blockchain ledger. The hash, a short sequence of bits, is derived from a combination of the existing ledger and the block that is to be added; this alters whenever the contents of the entry are changed. Again, it is relatively easy to find the hash of a block (to process information to add a record) but difficult to pick a block that would yield

a specific hash value. That would require reversing the process to derive the information that generated the hash.

Bitcoin also requires that the hash meets a mathematical condition. Anyone who wishes to add a block to the ledger must keep their computer running a random search until that condition is reached. This process slows the addition of blocks, giving time for everything to be recorded and checked by everyone in the network. It also stops any individual from monopolizing network administration, because anyone with sufficient computational power can contribute blocks.

Yet, within ten years, quantum computers will be able to calculate the one-way functions, including blockchains, that are used to secure the Internet and financial transactions. Widely deployed one-way encryption will instantly become obsolete.

Information security has faced such mass extinctions before. For example, during the Second World War, German military messages were encoded and decrypted using Enigma machines, initially giving the Axis powers an advantage until the Allies cracked the Enigma code. And in 1997, the Data Encryption Standard, an algorithm

"A wrongdoer equipped with a quantum computer could forge any digital signature."

for encrypting electronic data that was then state of the art, was broken in a public contest to prove its lack of security. That gave rise to a second competition to develop a new protocol, resulting in today's Advanced Encryption Standard.

QUANTUM ADVANTAGE

Quantum computers exploit physical effects, such as superpositions of states and entanglement, to perform computational tasks. They are currently much less powerful than conventional computers, but will soon be able to outperform them on certain tasks. One such example is breaking security protocols that are based on cryptographic algorithms, as mathematician Peter Shor pointed out in 1994 (ref. 3). A blockchain is particularly at risk from this because one-way functions are its sole line of defence — a user's only protection is their digital signature, whereas bank clients are protected by plastic cards, security questions, identity checks and human cashiers.

Cracking of digital signatures is therefore the most imminent threat. A wrongdoer equipped with a quantum computer could use Shor's algorithm to forge any digital signature, impersonate that user and appropriate their digital assets. Most specialists think that this feat would require a universal quantum computer (one capable of performing a wide variety of calculations),



which is more than a decade away. Yet some researchers suggest that this could happen sooner, using emerging quantum computational devices that have more limited capabilities, such as those being developed by the computing firms D-Wave, Google and others^{4,5}.

Quantum computers will find solutions quickly, potentially enabling the few users who have them to censor transactions and to monopolize the addition of blocks to the Bitcoin ledger (known as mining). These parties could sabotage transactions, prevent their own from being recorded or double-spend. An international team of researchers has highlighted the possible impacts of such attacks⁶, with a report earlier this year charting the threats and suggesting a possible workaround⁷.

If nothing is done to update the protocols, cryptocurrencies will crash once quantum computers become available.

IMPROVING SECURITY

Fortunately, quantum technologies also offer opportunities to enhance the security and performance of blockchains.

Quantum-safe encryption. Quantum communications are inherently authenticated — no user can impersonate another. Such technologies use states of individual particles of light (photons) to encode bits and communicate them. Fundamental

physics stipulates that quantum states cannot be copied or measured without being altered. Any eavesdropper will be immediately uncovered.

Quantum cryptography can be used to replace classical digital signatures and to encrypt all peer-to-peer communications in the blockchain network. Our group has demonstrated such a simple system⁸. However, the complexity and cost of quantum cryptography networks will limit their adoption. In particular, current protocols require that each node in the network be connected to every other through optical fibre channels, because there is no trust in any intermediary node and hence all communications must be direct. Protocols will be needed to maintain secure communications even when information flows through untrustworthy nodes; these systems have been developed but need to be made more accessible for consumers.

Photon losses in optical fibres are another challenge. These limit the range of modern quantum-key distribution systems to a few tens of kilometres. The solution is to develop a quantum repeater, which uses quantum teleportation and quantum optical memory to distribute entangled states between the communicating parties. Research is progressing, but is a long way from delivering a practical device.

In the interim, one-way functions should be tightened. Some alternative encryption functions have been proposed⁹ that should be equally difficult to reverse using conventional or quantum computers. Although not completely secure, these could be run on existing hardware and would buy time, but they, too, could be deciphered in the long term.

Quantum internet. Using quantum technology for communicating as well as for the computational processing of blockchain data would further enhance security and enable blockchains to become faster and more efficient. This step requires a ‘quantum internet’¹⁰ — connecting quantum computers across a quantum communications network. It would then become possible to run fully quantum blockchains. These would bypass some computationally intensive steps of the current verification and consensus processes, and thus be more efficient and more secure. The proposed Quantum Bitcoin currency could be realized, with its security assured by the no-cloning theorem of quantum mechanics. Such quantum ‘bank notes’, if they still prove necessary in future, could be made impossible to forge by containing quantum information records¹¹.

The quantum internet is several decades away, so ‘blind quantum computation’ is an interim step. In this, a user with a conventional computer could run an algorithm on a remote quantum computer without sharing the input data or algorithm. This

technology would enable public cloud-quantum-computing platforms, making blockchains cheaper and more accessible.

NEXT STEPS

The blockchain business needs to update its existing software to use one-way cryptographic functions that are equally hard to reverse using conventional or quantum computers⁹. Until these post-quantum solutions are established or standardized, platforms must be flexible and capable of changing cryptographic algorithms on the fly¹².

The longer-term answer is to develop and scale up the quantum communication network and, subsequently, the quantum internet. This will take major investments from governments. However, countries will benefit from the greater security offered¹³. For example, Canada keeps its census data secret for 92 years, a term that only quantum cryptography can assure. Government agencies could use quantum-secured blockchain platforms to protect citizens’ personal financial and health data. Countries leading major research efforts in quantum technologies, such as China, the United States and members of the European Union, will be among the early adopters. They should invest immediately in research. Blockchains should be a case study for Europe’s Quantum Key Distribution Testbed programme, for example.

Much greater urgency needs to be given to these risks — their impact could be grave. ■

Aleksey K. Fedorov is a quantum information-technology group leader; **Evgeniy O. Kiktenko** is a leading research fellow; and **Alexander I. Lvovsky** is a quantum-optics group leader at the Russian Quantum Center, Moscow, Russia. **A. I. L.** is also a professor in the Department of Physics, University of Oxford, UK. e-mails: akf@rqc.ru; e.kiktenko@rqc.ru; alex.lvovsky@physics.ox.ac.uk

- Marr, B. ‘How Blockchain Technology Could Change The World.’ (*Forbes*, 27 May 2016).
- Nakamoto, S. *Bitcoin: A Peer-to-Peer Electronic Cash System* (Bitcoin, 2008).
- Shor, P. W. in *Proc. 35th Ann. Symp. Found. Comp. Sci.* 124–134 (IEEE Comp. Soc. Press, 1994).
- Peng, X. *et al. Phys. Rev. Lett.* **101**, 220405 (2008).
- Anschuetz, E. R., Olson, J. P., Aspuru-Guzik, A. & Cao, Y. Preprint at <https://arxiv.org/abs/1808.08927> (2018).
- Aggarwal, D., Brennen, G. K., Lee, T., Santha, M. & Tomamichel, M. Preprint at <https://arxiv.org/abs/1710.10377> (2017).
- Stewart, I. *et al. R. Soc. Open Sci.* **5**, 180410 (2018).
- Kiktenko, E. O. *et al. Quantum Sci. Technol.* **3**, 035004 (2018).
- Bernstein, D. J. & Lange, T. *Nature* **549**, 188–194 (2017).
- Kimble, H. J. *Nature* **453**, 1023–1030 (2008).
- Broadbent, A. & Schaffner, C. *Des. Codes Cryptogr.* **78**, 351–382 (2016).
- Gheorghiu, V., Gorbunov, S., Mosca, M. & Munson, B. *Quantum-Proofing the Blockchain* (Univ. Waterloo, 2017); available at <https://go.nature.com/2b2uvft>
- Chapron, G. *Nature* **545**, 403–405 (2017).



A river delta in Iceland, seen from the air.

CONSERVATION

Earth reframed: a big idea

Huw Lewis-Jones reflects on how conceptualizing nature anew became a call to action.

Are we robbing the next generation by impoverishing the planet? Can we find a way for economies to grow without depleting the environment? Barely 50 years ago, such provocative questions might have seemed unimaginable. This was not because people were unaware of the damage already wreaked. Instead, as an intriguing book reveals, no one had fully conceptualized the intricate interconnections of nature. Without that framing, humanity could not adequately describe the scale of its own impact on the planet. From the infinitely complicated was born a simple term: the environment.

In *The Environment: A History of the Idea*, environmental historians Paul Warde, Libby Robin and Sverker Sörlin trace the concept's emergence from 1948 to today. They show that in the years following the Second World War, awareness grew of humanity's capacity for cataclysmic destruction. Fears for the

future ignited a desire to improve definitions of Earth systems. The environment as a concept was nurtured over succeeding decades in political demonstrations, unsung conferences and drawn-out legislative processes. The story is also one of new tools of measurement and interdisciplinary thinking, the aggregation of scientific results and shifting authorities, later enabled and enhanced by the digital revolution.

A vocabulary evolved to frame human-driven effects on the planet during post-war reconstruction, a feverish period for



The Environment: A History of the Idea

PAUL WARDE, LIBBY ROBIN, AND SVERKER SÖRLIN
Johns Hopkins University Press
(2018)

building global institutions and philosophies. In 1948, one of the first large-scale environmental organizations — the International Union for the Protection of Nature (IUPN), later the International Union for Conservation of Nature — was founded in Fontainebleau, France. Its mission was the “preservation of the entire world biotic community”; the idea of placing absolute limits on essential resources, later captured in expressions such as ‘peak oil’, was born. Yet, as the authors point out, that year also saw the launch of Soviet leader Joseph Stalin’s ‘Great Plan for the Transformation of Nature’. In response to half a million deaths from drought and famine in 1946–47, Stalin ordered a programme of dam and irrigation construction to protect the future of agriculture on the steppes. It ultimately wreaked “new havoc, including the desiccation of the Aral Sea”.

‘Thinking globally’ had long predated

the IUPN. Nineteenth-century polymath Alexander von Humboldt, for instance, saw nature as a “living whole” and predicted climate change. Soviet scientist Vladimir Vernadsky introduced the idea of Earth’s life-supporting zone in his 1926 book *Biosfera*. But by the 1950s, an infrastructure of expertise and mainstream will had arisen to sustain the global approach.

A new cadre of experts was called upon to help inform public understanding. They included biologist Rachel Carson — who had been publishing popular books on marine biology since 1941 — and ecologist William Vogt, author of the 1948 *Road to Survival* (A. Rome *Nature* 553, 152–153; 2018). In 1954 came geochemist Harrison Brown’s rumination on planetary resources and human population, *The Challenge of Man’s Future*. In 1970, the first Earth Day took place — and systems scientist Jay Forrester developed a model of global dynamics based on 120 lines of computer code. That laid the basis for the Club of Rome’s hugely influential report *The Limits to Growth* in 1972.

The environment as an idea “burst into life in a futurological soup”, as the authors write, but it was also driven by pioneering scientists compelled to provide solutions to the degradation they witnessed. The ambitious 1955 international symposium *Man’s Role in Changing the Face of the Earth*, called by the New York-based Wenner-Gren Foundation for Anthropological Research, began to shift the emphasis to humanity as culprit. It brought together the likes of geographer Carl Sauer, zoologist Marston Bates and urban-planning theorist Lewis Mumford, but regrettably only one woman: plant geneticist Janaki Ammal, then leading the Botanical Survey of India.

A decade on, Future Environments of North America, convened by the US Conservation Foundation in Warrenton,

Virginia, and including many of the same people, looked to extend those ideas across disciplines, from conservation to geology, economics and sociology. ‘Public policy’ and ‘management’ became part of established discourse, and the vision of Canadian ecologist Pierre Dansereau was widely embraced: “A valid imaginary reconstruction of our world is now our greatest task. It may even be the condition of our survival.”

Five decades on, that warning is more important than ever. Overwhelming evidence reveals how Earth-system processes, from hydrology to biology, are altered by human activity. With the concept of the Anthropocene, an epoch defined by human impact on Earth, reaching the mainstream — as much metaphor as formal term — it is useful

to look back and consider how conversations sustaining this theme first found voice, and to examine the challenges this radical way of thinking faced. Many now well-known threads in modern conservation and ecology — resources, biodiversity, pollution and climate change — have a cultural history. *The Environment* maps that territory well.

As I was reading it, the 2018 special report of the Intergovernmental Panel on Climate Change launched in South Korea. It finds that limiting global warming to 1.5°C above pre-industrial levels, rather than 2°C, (as pledged in Paris in 2015) will require “rapid and far-reaching” transitions in land use, energy, industry, buildings, transport and cities. Refusal will trigger a staggering sequence of knock-on effects

“As environmental movements past have shown, we need imagination, accuracy, long-term political will — and hope.”

at every scale. At least twice as many key insect pollinators and plants would be likely to lose half their habitat. Corals would be 99% lost. There are thousands of other desperate scenarios, across species and landscapes. Beyond “integrated expertise” — a concept unfortunately challenged by leaders of some of the world’s most powerful economies — concerted, immediate action by governments is imperative. Individual action is still crucial, but might not be enough without policies that look beyond the political short term.

In 1987, Earth scientist Wallace Broecker noted in *Nature*: “We play Russian roulette with climate, hoping that the future will hold no unpleasant surprises” (W. S. Broecker *Nature* 328, 123–126; 1987). That narrative of ecological collapse has finally found its audience: as Warde, Robin and Sörlin show, we have progressed since the word environment bubbled into public consciousness. But some still refuse to listen. How can so central a concept retain urgency and impact in the decontextualized online war of words, truth, lies, expertise and its rejection by the march of populism?

As environmental movements past have shown, we need imagination, accuracy, long-term political will — and hope. “The environment is about people, too,” the authors note, “and how they respond to its changes and challenges.” Our relationship to nature goes far beyond resources, amenity or the scientific idea of an archive we learn to read. There are, as *The Environment* shows, ethical complexities in how we use and abuse the planet — and in how we frame its improbable riches. ■

Huw Lewis-Jones is an environmental historian, expedition guide and senior lecturer at Falmouth University, UK. Twitter: @polarworld

SPACE SCIENCE

A rock legend retells the race to the Moon — in 3D

Queen guitarist and astrophysicist Brian May’s latest collaboration is a stereoscopic delight, finds **May Chiao**

In 2019, it will be 50 years since the first Moon landing. Almost more remarkable is that no human has touched that surface since Gene Cernan’s lunar stroll during the last of NASA’s Apollo missions in 1972. To celebrate the scale of that programme, writer and editor David Eicher, together with

Brian May — astrophysicist, Queen guitarist and stereoscopic photographer — take us back to the beginning in the spectacular *Mission Moon 3-D*.

On 25 May 1961, President John F. Kennedy declared that the United States would land a man on the Moon and bring him

Mission Moon 3-D: Reliving the Great Space Race
DAVID J. EICHER AND BRIAN MAY
London Stereoscopic Company (2018)

safely back before the decade was out. That ambitious timetable surprised the president’s own science advisers, as well as the US Congress and the rest of the world. Kennedy was spurred by the phenomenal successes of Soviet space exploration. Sputnik 1, launched in 1957, was the first artificial satellite to orbit Earth; and a month before Kennedy issued his challenge, Yuri Gagarin had become the first human in space.

The story of how NASA overtook the Soyuz programme during the cold war has been told many times — in interviews and books and on film. So what do Eicher and May bring to the table?

Primarily, Eicher compares the Soviet ▶

► and US space programmes — their successes and failures. This political, cultural and technical context is enriched with information that has come from the cosmonauts themselves in recent years. For example, details of the accidents and deaths that hindered the Soviet lunar programme, from the cosmonauts' point of view, enable Eicher to tell a more complete story. He strikes a fine balance between detail and readability.

But the book is so much more. Its 150 stereo photographs, which can be seen in 3D through a stereo viewer, make it an immersive experience. Since childhood, May has collected stereoscopic devices — a Victorian technology in which two photographs of the same subject (taken a small horizontal distance apart) are displayed side by side. Looking at these through a viewing device, at a certain distance and with eyes 'relaxed', the brain creates the perception of depth, and previously unresolved details jump into focus. The pairs of images that Eicher and May include show everything from cosmonaut Alexei Leonov, the first spacewalker, in 1965, to the *Apollo 12* lunar module *Intrepid* flying insect-like above the Moon's surface in 1969. A hand-held LITE OWL viewer developed by May is included with the book with instructions (see go.nature.com/2ezgyg6). For those struggling to see in 3D, try starting with high-contrast images such as the one of Comet 67P/Churyumov-Gerasimenko.

Stereo photography was not an aim of the Apollo missions. But many sequential photographs were taken — for instance by Stuart Roosa in *Apollo 14* while circling the Moon — which enabled May to assemble several pairs. May and his team also trawled the NASA archives to find serendipitous pairs of photographs or film stills with just the right baseline separation. To illustrate the



A stereoscopic image of US astronaut Gene Cernan next to a lunar rover during an Apollo 17 moonwalk.

Soviet effort, for which no sequential images existed, they had to convert 'mono' photographs into stereo pairs.

As these vivid images remind us, the pace of progress would have been much slower without the fierce competitiveness of the space race. However, the cold-war wall between the two countries made avoidable, sometimes tragic, mistakes inevitable. One chilling example is the *Apollo 1* accident in 1967. During a routine countdown rehearsal, a fire erupted in the craft's main capsule, which contained pure oxygen; astronauts Roger Chaffee, Gus Grissom and Ed White died almost instantly. (Only later was a quick-release hatch added to the design.) Six years before, unbeknown to NASA, trainee cosmonaut Valentin Bondarenko had suffered a similar fate during a test in Moscow. The two superpowers' first cooperative space-flight would have to wait until 1975.

Mission Moon 3-D devotes significant

space to the ultimate sacrifice made by humans (and animals) in the name of space exploration, underlining the risks of propelling earthlings into an alien environment. Now, NASA, the Russian, Japanese and Chinese space agencies, and the private companies SpaceX and Blue Origin, plan to send humans back to the Moon. Before that happens, any benefits must be weighed carefully against the risks, and the expense. Reaching Mars will demand that several nations work together, with involvement from the public and private sectors. Robotic and telescopic missions cost much less and can reach more-distant planets and moons. But there is no substitute for human experience; and while we wait for another foot to fall on an extraterrestrial landscape, books such as this one give us an inkling of that ultimate thrill. ■

May Chiao is chief editor of *Nature Astronomy*.

TECHNOLOGY

The *Doctor Who* theme and beyond: female pioneers of electronic music

Joanne Baker lauds a paean to the experimentalists of the BBC Radiophonic Workshop.

The history of electronic music usually centres on the men (including Pierre Schaeffer, Olivier Messiaen, Pierre Boulez, Karlheinz Stockhausen and Edgard Varèse) who developed *musique concrète* from recorded everyday sounds in Paris in the mid-twentieth century. Also in those decades, a group of sound engineers — many of them women — were making waves in an old London skating rink.

The BBC Radiophonic Workshop

Synth Remix

93 Feet East, London.

8 November 2018; Touring 8–11 November.

produced effects and theme tunes for the British broadcaster, including iconic sounds for the sci-fi television and radio programmes *Doctor Who* and *The Hitchhiker's Guide to the Galaxy*, using electronic oscillators and tape loops decades before synthesizers were common. That many of

its engineers were women was, and still is, a rarity. Last week, two of them, Daphne Oram and Delia Derbyshire, were celebrated anew in Synth Remix, a concert series of live performances and DJ sets touring Britain.

Oram (1925–2003) co-founded the Radiophonic Workshop. She gained experience in mixing electronics and music during the Second World War while working for the BBC on sound balance for radio broadcasts. During Germany's bombings of London in

the Blitz, she switched pre-recorded tracks of orchestral music into broadcasts of live music. That allowed the musicians to flee the city's grand concert venue, the Albert Hall, without the radio audience knowing.

In the 1950s, Oram became intrigued by the potential of tape recording to transform music by exploding space and time. She was a fan of *musique concrète*, regularly staying up all night to mix her own tracks. In 1958, after years of badgering the BBC to modernize its music, Oram and her colleague Desmond Briscoe were given a room with some old equipment. Thus began the workshop.

Oram left after just a year. The BBC asked her to take six months off, saying it was concerned that the equipment might have adverse effects on the human body. So she quit.

Oram set up her own home studio in a converted rural oast house in Kent. She continued to compose electronic sounds, and to lecture and write about the nature of vibrations. She launched a field that she called Oramics, using a device that she built for 'drawing' sound. The size of a dressing table, it subverted the technology behind the cathode-ray oscilloscope, which converts sound waves into a picture. Lines, squiggles and dots sketched on 35-mm film were scanned and used as indications of pitch, vibrato and timbre. (It was, in effect, an early sequencer — a technology that eventually came along in the 1980s.) In her seminal 1972 book *An Individual Note of Music, Sound and Electronics*, she wrote of humans as instruments, harbouring "a whole spectrum of resonant frequencies" that are "vibrant with pulsating tension".

Oram paved the way for Derbyshire (1937–2001), who famously crafted the unearthly 'sweeps and swoops' of the *Doctor Who* theme

tune in 1963. Derbyshire told interviewers that her love of abstract sounds came from the air-raid sirens she heard growing up in Coventry during the Blitz, recalling that "the sound of the 'all clear'" was electronic music. She studied mathematics and music at the University of Cambridge, and took an analytical approach to experimenting with sound. Her notes, now archived at the University of Manchester, are full of mathematical symbols and equations. She jotted down explicit frequencies and used the dots and dashes of Morse code.

Her sketched scores are visual — crescendos of squiggles, rings of organic contours and hatched textures of mass and void. Triangular bursts march across the page like streams from flak guns. Some rounded forms look like the Lissajous patterns (formed from interacting sine waves) that she must have seen on oscilloscope displays.

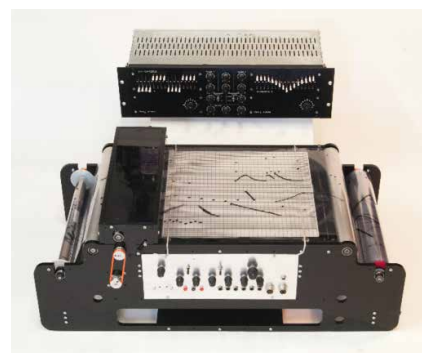
"Her sketched scores are visual — crescendos of squiggles, rings of organic contours and hatched textures of mass and void."

Her notes are also littered with evocative words: hum, beat, splash, shimmer.

This Radio-phonetic legacy was the launching point for the concerts, performed by musician Jo Thomas and artist

Olivia Louvel. At the first performance at 93 Feet East, a concrete venue in east London, each deftly controlled networks of tabletop electronics. These sent tsunamis of sound waves shuddering through the audience's chests, and lofted clouds of siren vocals around our heads. The compositions were compelling, richly textured and darkly powerful.

Electronic music integrates science with



An Oramics machine for drawing sound.

art, and Thomas has roamed far into that liminal space. In 2012, her *Crystal Sounds of a Synchrotron* — based on blips and beeps from the Diamond Light Source particle accelerator near Didcot, UK — won an international award for digital music and sound art from Austrian arts body Ars Electronica. Here, her three-part work *Nature's Numbers* nods to Derbyshire and Oram using a bank of self-built electronic components, conventional synthesizers and computers augmented with tones of her own voice. Thomas immersed herself in Derbyshire's archives for Synth Remix.

'Time Lament', the first part of Thomas's performance, combines high, plaintive vocals over a heartbeat reminiscent of spinning tape loops. Inspirations include Derbyshire's love of seventeenth-century composer Henry Purcell (and his aria 'Dido's Lament') and effects from *The Hitchhiker's Guide to the Galaxy*. In 'Echoes of the Earth', sounds like showers of rain punctuate the eerie silence of a cavern, as a homage to Derbyshire's vocal experimentations. 'Cellophane Resonance' is a playful collage of sci-fi sound effects. Here, Thomas exploits a compact reproduction of an Oramics machine, generating sounds from scribbles on what look like rolls of transparent film from an overhead projector.

Louvel's multimedia suite *Data Regina* liberates the voice of another woman from the past. Tudor monarch Mary, Queen of Scots — ultimately executed by her cousin, Elizabeth I of England — was a prolific writer and poet. Yet few know her works. Louvel's voice, computer music and a video backdrop of avatars that look like chess pieces transmit the story of the tortured queen. The result is a soaring, gut-wrenching opera.

These intuitive and democratic takes on compositions by Oram and Derbyshire reverberate today. You don't need a bank of high-powered electronics to pick out a beat or strike a chord. Go on, grab your laptop, switch on an app and play with sound. This, in essence, is what they did. ■

Joanne Baker, a senior Comment editor at *Nature in London*, has written three books about physics.



Daphne Oram in her home studio in 1962.

Correspondence

Support for African clinician scientists

As members of Africa's next generation of scientists, we agree that there is a need to build the capacity of African institutions to train skilled researchers and scholars (see go.nature.com/2araify). To this end, we recently founded the South African Clinician Scientists Society (www.sacss.co.za). By creating a collegial environment for emerging specialists, we hope this, and similar societies, will contribute to attracting and retaining African scientists and easing their scientific journey (see *Nature* **562**, S58–S61; 2018).

Researchers who return from training abroad to a supportive and enabling environment make the transition into successful independent scientists faster than do isolated researchers.

So, the society uses three strategies to nurture talented professionals, centred on relationships. First, it identifies suitable mentors. Second, the society develops research networks. Third, it aims to create multidisciplinary research units that provide administrative and research support.

Salome Maswime *Massachusetts General Hospital, Boston, Massachusetts, USA.*

Gwinyai Masukume *Irish Centre for Fetal and Neonatal Translational Research, Cork, Ireland.*

Nomathemba Chandiwana *Wits Reproductive Health and HIV Institute, Johannesburg, South Africa.*
smaswime@mh.harvard.edu

Equity more likely in diverse labs

The unethical exploitation of others' work, particularly by individuals who benefit from social privilege, is all too common in the current 'winner takes all' model of research (see *Nature* **553**, 367–369; 2018). In my view, workforce diversity can help. Privilege and

entitlement are self-propagating phenomena that thrive on overrepresentation. Their effects are tempered when the privileged are not in the majority.

In my experience, equity prevails in groups with no apparent ethnic or gender bias. I have been a student in six labs and collaborated with countless others. My current lab is the most heterogeneous: of 37 members, 14 of us are female and our backgrounds span 17 countries. Only two fit the description of 'white American male'. I find it empowering to work in a gender-diverse, multicultural environment that is quick to rebuke entitled behaviour.

Of course, diversity alone is no panacea. Skewed power dynamics are an almost inevitable consequence of the financial and reputational dependence of postdoctoral fellows and graduate students on professors, as well as of the hypercompetitive nature of scientific research (see *Nature* **533**, 452–454; 2016). But for investigators putting together their own research groups, engineering diversity is a productive first step towards a more humane science system.

Amin Aalipour *Stanford University School of Medicine, Stanford, California, USA.*
aalipour@stanford.edu

Safeguard our audiovisual heritage

The fire at Brazil's National Museum in Rio de Janeiro in September destroyed many audiovisual recordings, including some of extinct South American Indigenous languages. This is an immeasurable loss to our record of biological diversity and worldwide culture. We urge the scientific community to deposit and digitize recordings in institutional archives, then to replicate and store them to guard against any future damage.

Audiovisual collections preserve human history, allow

population monitoring and provide insight into animal natural history. In fields such as taxonomy, diversity and conservation, photos and videos, for example, might be the only way of ensuring species diagnosis for specimens that deteriorate soon after preservation.

Yet scientists can be lax about archiving. For example, only 22% of South American herpetologists have uploaded their amphibian recordings to a shared repository (R. R., unpublished). Some resist doing so because they do not want their data to be publicly available.

Depositing and digitizing analogue media are not enough to safeguard our audiovisual legacy. It is essential to back up deposited media and use cloud-based storage as well.

Luís Felipe Toledo* *Unicamp, Campinas, Brazil.*

**On behalf of 4 correspondents (see go.nature.com/2b8r8xc for complete list).*
toledosapo@gmail.com

Join forces to tackle antibiotic resistance

Shortly after it was revealed that important antibiotics are being used in "unacceptable" quantities on US farms (see go.nature.com/2zstks6), reports surfaced that the United Kingdom might not permanently commit to European Union plans to limit use of the drugs in agriculture (see go.nature.com/2an1r4q). Loosening regulations to facilitate trade might seem attractive, but it could weaken the only existing transnational antibiotic stewardship coalition.

History shows that global collective action is necessary to tackle antimicrobial resistance (AMR). Since the 1940s, physicians have reported AMR across the world. In 1954, the first 'superbug' — *Staphylococcus aureus* 80/81 — swept around the planet. Knowledge of AMR's border-defying hazards

failed to trigger coordinated responses. Scandinavians restricted medical prescriptions; Americans opted for educational measures. In agriculture, Germans targeted antibiotic residues in food and the United Kingdom restricted medically relevant antibiotic growth promoters (AGPs). Fragmented policies created the current patchwork of antibiotic use (see C. Kirchhelle *Palgrave Commun.* **4**, 96; 2018).

Patchwork regulations won't work. Take agriculture: the same products are used in medicine and farming, but can be subject to different rules. Overuse in one sector can undermine restrictions in another. The EU banned AGPs in 2006, but initially failed to regulate veterinary surgeons supplying the same drugs for prophylactic or therapeutic purposes. Narrow AGP restrictions and toothless enforcement now also haunt US regulators. Meanwhile, global antibiotic-production centres (such as India and China) no longer align with Western centres of policing.

Promoting international surveillance of AMR and drug use can remedy fragmented policies. In high-income nations, it has aided research and stewardship. These countries have a responsibility to share the burden of stewardship and support poorer countries to improve theirs. National efforts will achieve little on their own.

Claas Kirchhelle *Oxford Martin School, University of Oxford, Oxford, UK.*

claas.kirchhelle@wuhmo.ox.ac.uk

CORRECTION

The Outlook article 'Expanding the reach of science' (*Nature* **562**, S10–S11; 2018) cited the wrong value for the number of STEM teachers in Accra who have been trained by The Exploratory. It should have been 70, not 700.

Paul G. Allen

(1953–2018)

Microsoft co-founder who established the Allen Institute for Brain Science.

The world knew technology billionaire Paul G. Allen as the other founder of Microsoft — Bill Gates's erstwhile partner in revolutionizing personal computing. Sports fans knew him as the owner of Super-Bowl-winning football team the Seattle Seahawks. To scientists, he was the philanthropist behind the Allen Institute, known for its pioneering brain-mapping research and cell science, and the Allen Institute for Artificial Intelligence.

Allen, who died on 15 October, was born on 21 January 1953 in Seattle, Washington, a city to which he remained faithful throughout his life. With school friend Bill Gates, in 1975 he founded Micro-Soft, purveyors of operating systems for the nascent desktop computer market. Bereft of its original hyphen, it grew into one of the world's most valuable companies, netting Allen a vast personal fortune.

In 1983 he withdrew from day-to-day involvement in the company to deal with early stage Hodgkin's lymphoma (although he remained on the board until 2000). This close encounter with death and his wealth freed him to pursue passions seeded in his youth — music, sports, the environment, science fiction, space travel and science.

Paul was attracted to the vast complexity of biology. He was intrigued by how the 3.2 billion nucleotide letters arrayed as strands of DNA in a single fertilized egg give rise to the 30 trillion cells that make up a human. In March 2002, after the success of the Human Genome Project, Paul convened meetings with geneticist James Watson and others to focus on a big biology project of his own.

Paul wanted deliverables and milestones. He had been burned by an attempt in the 1990s to seed innovation at his technology incubator Interval Research in Palo Alto, California, where he hired talents from Stanford University, the Massachusetts Institute of Technology and Bell Labs, and gave them carte blanche to work on Internet-related ideas. On the initiative of neurobiologist David Anderson at the California Institute of Technology, Paul and his sister Jody Allen started the Allen Brain Atlas in 2003 at the Allen Institute for Brain Science in Seattle. With the leadership of cell biologist Allan Jones, the project was delivered on time and under budget in 2006, and yielded a map of the spatial expression of 20,000 genes throughout the entire brain of the adult laboratory mouse, in a highly reproducible 3D framework.

The Allen Brain Atlas fulfils what biologist



Sydney Brenner calls the CAP criteria — a community resource that is complete, accurate and permanent. Its 3D coordinate system has become the pole star by which thousands of labs working on the mouse brain orient themselves. The mouse atlas now gets hundreds of thousands of page visits a year, twice as many as when it was created.

In quick succession, the institute produced gene-expression snapshots of the developing and mature brains of mice, non-human primates and humans. At Paul's insistence, and unusually for the time, all data, metadata and methodological white papers, were, and continue to be, freely and publicly available before associated discoveries are published. This has had a transformative effect on the field, and is now mandated by many funders.

Flushed with success, Paul was emboldened to ask harder questions concerning how the 100 billion neurons in the human brain give rise to intelligence, vision and action. He thought about this in terms of coding and programming. What is the code used for perception? Can our cognitive abilities — visual perception, short- and long-term memory, planning, reasoning, imagination, language and so on — be conceived of as applets running on the highly specialized hardware of the brain? What can theories of cortical computation teach us about the brain? Can we engineer cortical circuits in a dish? What is the difference between natural and artificial intelligence? (Paul started the Allen

Institute for Artificial Intelligence in 2014.)

I was recruited in 2011 to be the chief scientist for the second decade of his brain institute. Fired up by having survived a second bout of lymphoma in 2009, Paul tripled the institute's size and budget. He tasked us to carry out a census of all cell types in the mouse (see B. Tasic *et al.* *Nature* <https://doi.org/10.1038/s41586-018-0654-5>; 2018) and in the thousand-fold-bigger human brain, and to build the Allen Brain Observatory (see C. Koch and R. C. Reid *Nature* **483**, 397–398; 2012). This observatory is dedicated to large-scale surveys (similar to those in astronomy) of cellular level activity in mice, conducted using optical fluorescent microscopy and high-density electrical recordings.

In 2014, Paul started the Allen Institute for Cell Science, focused on visualizing the organelles inside engineered human cardiac cells. In 2015, the brain-mapping and cell-science efforts were amalgamated into a single Allen Institute, with more than 500 staff in a sleek new building in Seattle, led by Allan Jones.

Paul wasn't a scientist and didn't aspire to be. He was a gifted and intensely curious outsider who kept asking hard questions. His way of fuelling discovery was to empower entrepreneurial teams of scientists, engineers and staff, challenging them to draw up tangible, time-stamped goals and milestones. He was keen on knowing the answers, and kept pushing for them: "If not yet, why not? What is holding us up?" He asked us to make hard choices if the stated goals could not be achieved, including shutting down underperforming research programmes.

In person, Paul was discreet, almost diffident. At scientific advisory board meetings, he sat quietly in the back until asking the one critical question that galvanized the room and changed the project's trajectory (without ever presuming to know the answers; he was too humble for that).

At the time of his death from the same cancer he'd weathered in 2009, he was considering how his unique brand of mission-oriented, team-based science could answer some of biology's most persistent mysteries — in evolution, development, neuroscience, immunology, health and ecology. As captured in the title of his 2011 autobiography, Paul was ever the Idea Man. ■

Christof Koch is president and chief scientist of the Allen Institute for Brain Science in Seattle, Washington, USA.
e-mail: christofk@alleninstitute.org

ENGINEERING

Flying with ionic wind

Aeroplanes use propellers and turbines, and are typically powered by fossil-fuel combustion. An alternative method of propelling planes has been demonstrated that does not require moving parts or combustion. [SEE LETTER P.532](#)

FRANCK PLOURABOÛÉ

Small, lightweight devices called lifters can propel themselves into the air without combustion or moving parts, and have become a popular topic of discussion with technology buffs on social media in the past few years. And yet the physical mechanism behind lifters has been known for more than a century¹. When charged molecules in the air are subjected to an electric field, they are accelerated. And when these charged molecules collide with neutral ones, they transfer part of their momentum, leading to air movement known as an ionic wind. On page 532, Xu *et al.*² demonstrate that an aeroplane with a 5-metre wingspan can sustain steady-level flight using ionic-wind propulsion. Improvements are required, but the authors' proof-of-concept demonstration could pave the way for the development of enhanced propulsion systems.

In Xu and colleagues' plane, an electric field is applied to the region that surrounds a fine wire called the emitter (Fig. 1a). The field is strong enough to induce a chain reaction: free electrons in the region collide heavily enough with air molecules to ionize them, producing more electrons that then ionize more molecules. These electron cascades give rise to charged air molecules in the vicinity of the emitter — a phenomenon called a corona discharge. Finally, the charged molecules drift away from the emitter and generate a propulsive ionic wind as they are accelerated by the electric field towards a device called the collector (Fig. 1b). This process occurs only in gases, and not in liquids, justifying the authors' use of the term 'electroaerodynamics'.

Previous experiments suggested that ionic-wind propulsion could enable the steady-level flight of an aircraft, but that the feasibility of achieving this lies at the limit of what is currently technologically possible³. Xu *et al.* therefore needed to systematically search through all of the possible aeroplane designs for a feasible option. They used a technique called geometric programming to find the optimum set of design variables that would also minimize the aircraft's wingspan and, in turn, its weight, electrical-power requirements and cost.

The optimization technique found a feasible design at a wingspan of 5 metres, with a mass of

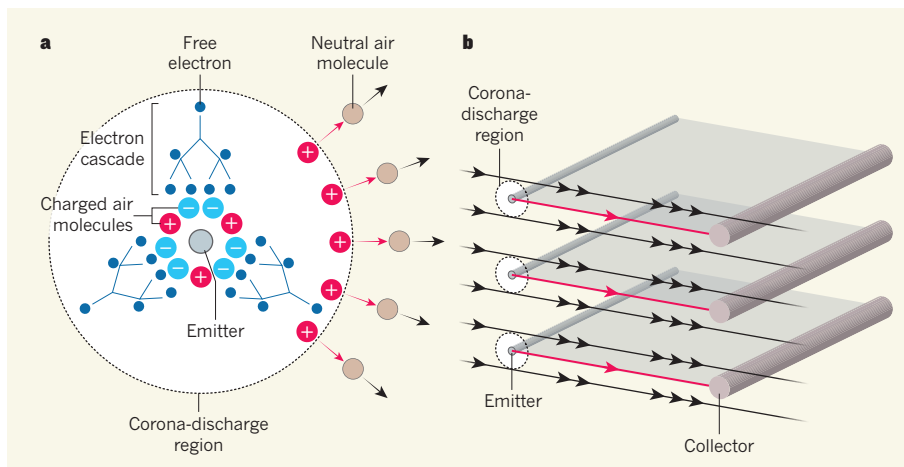


Figure 1 | Ionic-wind propulsion. Xu *et al.*² demonstrate that an aeroplane can sustain steady-level flight using air movement known as an ionic wind. **a**, In the authors' aircraft, an electric field (not shown) is applied to the region surrounding a fine wire called the emitter (shown in cross-section). The field induces electron cascades, whereby free electrons collide with air molecules (not shown in the cascades) and consequently free up more electrons. This process produces charged air molecules in the vicinity of the emitter — a corona discharge. Depending on the electric field, negatively or positively charged molecules drift away (red arrows) from the emitter. These molecules collide with neutral air molecules, generating an ionic wind (black arrows). **b**, The aircraft uses a series of emitters and devices called collectors, the longitudinal directions of which are perpendicular to the ionic wind. The flow of charged air molecules occurs mainly along the directions (red arrows) joining emitters and collectors. Consequently, the ionic wind is accelerated (black arrows) predominantly in these regions.

2.5 kilograms, a flight velocity of 4.8 metres per second and a power requirement of 600 watts. The authors built a full-scale plane based on this design (see Fig. 1b of the paper²). They flew the aircraft ten times, and showed that it achieved steady-level flight.

In the 1960s, various studies^{4,5} seemed to sound the death knell for propulsion based on ionic wind. They demonstrated that only about 1% of the input electrical energy was used in propulsion — not far from the 2.6% reported by Xu and colleagues. However, at least three factors make the approach appealing for aircraft.

First, it is now known that the energy efficiency improves substantially when the aircraft velocity is increased. For example, if the velocity reaches 300 m s⁻¹, the efficiency^{2,6} can be as high as 50%. Second, many studies have shown that ionic wind can enhance the aerodynamics of plane wings⁷. Third, the technique could facilitate what is known as distributed propulsion⁸, which is considered a major direction for improvement in aviation.

Aircraft propulsion is quantified by the freestream mass-flow rate — the total mass of air that passes through a given area in a given time. This rate is directly proportional to the cross-sectional area of the propulsion system, and to the increase in air velocity provided by the system. In distributed propulsion, an array of propulsion systems is spread along the length of the aircraft. This increases the total cross-sectional area and, in turn, the freestream mass-flow rate. But it also enhances the aerodynamic drag (the frictional force between the aircraft and the air). Using fine wires as the propulsion system, as Xu *et al.* did, could allow the total cross-sectional area to be greatly increased, while having almost no impact on the aerodynamic drag.

The scalability of the authors' propulsion system remains to be seen. Can ionic-wind propulsion fly an aircraft of several tonnes? This practical issue is still open, but predictions suggest that aircraft such as the solar-powered plane Solar Impulse 2 could sustain steady-level flight using only ionic wind⁹. An advantage of

ionic-wind propulsion systems, as opposed to propellers, is that they can be interfaced directly with batteries — the energy-storage devices of future planes — without affecting the rate of energy conversion. In the decades to come, drones or aircraft that use ionic wind might include secondary ionic-wind propulsion systems dedicated to energy saving and potentially coupled with solar panels.

These technological developments should provide a better understanding of the coupled physics of charged-molecule production and the resulting ionic wind that is central to such

propulsion systems. The force generated by ionic wind is directly proportional to the electric current that flows in the system^{2,10}. Because this current is strongly dependent on the configuration of emitters and collectors, research into the conception and optimization of ionic-wind propulsion can now begin, thanks to the breakthrough by Xu and colleagues. ■

Franck Plouraboué is at the *Institute of Fluid Mechanics of Toulouse, Toulouse University, CNRS, INPT, UPS, 31400 Toulouse, France.*
e-mail: franck.plouraboue@imft.fr

1. Chattock, A. P. *Phil. Mag.* **48**, 401–420 (1899).
2. Xu, H. *et al. Nature* **563**, 532–535 (2018).
3. Gilmore, C. K. & Barrett, S. R. H. *Proc. R. Soc. Lond. A* **471**, 20140912 (2015).
4. Robinson, R. *Am. Inst. Elect. Engineers Pt I* **80**, 143–150 (1961).
5. Stuetzer, O. M. *Phys. Fluids* **5**, 534–544 (1962).
6. Bondar, H. & Bastien, F. *J. Phys. D* **19**, 1657–1663 (2000).
7. Kriegseis, J., Simon, B. & Grundmann, S. *Appl. Mech. Rev.* **68**, 020802 (2016).
8. Sehra, A. K. & Whitlow, W. *Prog. Aerosp. Sci.* **40**, 199–235 (2004).
9. Monrolin, N., Plouraboué, F. & Praud, O. *AIAA J.* **55**, 4296–4305 (2017).
10. Monrolin, N., Praud, O. & Plouraboué, F. *Phys. Rev. Fluids* **3**, 063701 (2018).

NEURODEGENERATION

Disease protein muscles out of the nucleus

Protein aggregation is a characteristic of several neurodegenerative diseases. But disease-associated aggregates of the protein TDP-43 have now been shown to have a beneficial role in healthy muscle. SEE ARTICLE P.508

LINDSAY A. BECKER & AARON D. GITLER

Most neurodegenerative disorders are characterized by the build-up of clumps of proteins in the brain¹. A prevailing view in the field is that these large protein assemblies are inherently abnormal and are toxic to cells. Vogler *et al.*² challenge this canon by reporting on page 508 that muscle cells can contain physiological, reversible protein aggregates that have features similar to the aggregates seen in neurodegenerative disease, but that actually seem to be beneficial.

The protein TDP-43 forms aggregates in nerve cells in nearly all cases of the neurodegenerative disorder amyotrophic lateral sclerosis (ALS, also known as motor neuron disease)³. TDP-43 aggregation is also seen in other diseases, including frontotemporal dementia (FTD)⁴ and inclusion body myopathy (IBM)⁵, in which neurons and muscle cells, respectively, degenerate. FTD and IBM share genetic risk factors with ALS, indicating that the three have common disease mechanisms. In each disease, aggregates of TDP-43 are specifically found in the cytoplasm of dying cells. TDP-43 also has a normal job in the nucleus of healthy cells, where it acts as an RNA-binding protein⁴.

Vogler *et al.* set out to investigate the behaviour of TDP-43 in healthy muscle. In doing so, they made a surprising observation. As expected, TDP-43 was located in the nucleus of muscle stem cells. But when the authors coaxed these cells to differentiate into young muscle fibres called myotubes, or if they used a chemical to injure a mouse's leg muscle to stimulate muscle regeneration,

TDP-43 accumulated in the cytoplasm. There, it formed transient granular structures, which the researchers dubbed myo-granules, before moving back to the nucleus a few days later, as the myotubes became mature muscle fibres (Fig. 1). These data suggest that cytoplasmic TDP-43 myo-granules could have a role in muscle formation and regeneration.

Do myo-granules resemble the TDP-43 aggregates associated with neurodegenerative diseases? Disease aggregates are typically held together by strong bonds that are resistant to even heavy-duty detergents. Likewise, Vogler and colleagues found that TDP-43 myo-granules were resistant to

such detergents. Another key feature of many neurodegenerative-disease proteins (although not all disease-associated TDP-43 aggregates) is that they can adopt a specific conformation, known as amyloid. Amyloids are long fibres made up of building blocks of the misfolded disease proteins arranged in a highly organized manner⁶. Using an array of analytical methods — including an antibody to specifically detect amyloid-like material, and high-resolution microscopy and X-ray diffraction techniques to enable examination of the myo-granule's structure — the authors demonstrated that TDP-43 myo-granules have amyloid-like properties.

Next, Vogler *et al.* investigated differences between TDP-43 in cytoplasmic myo-granules and in the nucleus, by examining the RNAs to which the protein binds in the two settings. They found that the types of messenger RNA that bind to TDP-43 changed markedly as muscle precursors differentiated into muscles. The mRNAs found associated with aggregated TDP-43 included those that encode proteins associated with the sarcomere — a unit of muscle structure that causes muscle contraction. These data suggest that TDP-43 myo-granules might control the development of sarcomeres.

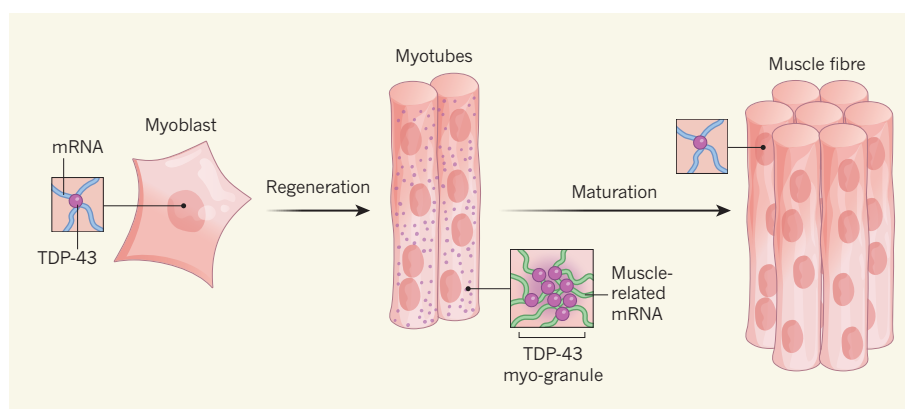


Figure 1 | A functional aggregate forms during muscle regeneration. In muscle precursor cells called myoblasts, the protein TDP-43, which binds messenger RNA, is located in the nucleus. Following muscle injury, myoblasts fuse into multi-nucleated fibres called myotubes that mature into muscle. Vogler *et al.*² show that TDP-43 transiently leaves the nucleus and assembles into large aggregate structures dubbed myo-granules, in which the protein binds to, and so might regulate, a distinct set of mRNA molecules involved in muscle formation. After recovery from injury, as the muscle matures, the myo-granules disassemble and TDP-43 returns to the nucleus.

To confirm a role for TDP-43 in muscle formation, the authors generated mice whose muscle stem cells lacked one of two copies of the gene that encodes the protein. Lowering the level of TDP-43 in this way led to a decrease in the diameter of the muscle fibres generated in response to injury, indicating that TDP-43 is important for full muscle regeneration — probably because it somehow regulates the expression of muscle mRNAs. However, this experiment does not prove that myo-granule formation is necessary for TDP-43 function in muscle regeneration; reducing TDP-43 levels causes cellular dysfunction in many cell types, but Vogler *et al.* report myo-granules only in myotubes.

Regardless of the physiological function of TDP-43 myo-granules, the authors' data beg the question of whether these structures can eventually turn into disease aggregates. To investigate this possibility, the group turned to mice carrying a mutated form of the gene *VCP* that can cause ALS, FTD and IBM in humans⁷. The mutant mice, in which muscle, brain and bone tissue degenerates⁵, had many more myotubes harbouring TDP-43 myo-granules than did wild-type mice. This suggests that *VCP* mutations might increase the risk of tissue degeneration by increasing the prevalence of myo-granules. In this scenario, perhaps small seeds of TDP-43 from myo-granules could be transported to the nerves that innervate muscle, where they might initiate a cascade of TDP-43 aggregation. Indeed, the earliest signs of neurodegeneration in ALS seem to originate at the nerve terminals adjacent to muscle, resulting in a 'dying-back' phenomenon that eventually reaches the main body of the neuron, which houses the nucleus⁸.

The differences between TDP-43 disease aggregates and myo-granules are as interesting as the similarities. Unlike myo-granules, most TDP-43 disease aggregates seem to have an amorphous structure, although some do have amyloid-like characteristics⁹. Moreover, the disease aggregates seem to be irreversible, whereas myo-granules disassemble as muscle cells mature. Because of this, myo-granules could provide an opportunity to investigate how strongly bound aggregate structures are disassembled. Factors that promote the disassembly of myo-granules might also be effective at clearing disease-associated aggregates.

Vogler and colleagues' findings raise an intriguing question. Strenuous exercise and weight training stimulate repeated rounds of muscle growth and repair — could this activity increase the production of TDP-43 myo-granules, increasing the propensity of TDP-43 to aggregate and so leading to diseases such as ALS? Indeed, there is some evidence for increased prevalence of ALS in elite athletes^{10,11}. However, much more evidence for the role of myo-granules and more human data will be needed before such a link can be assumed.

This paper sets the stage for future work

characterizing the physiological function and regulation of TDP-43 myo-granules, and for investigating how these complexes might contribute to disease. There are other examples of amyloid-like protein complexes that form in healthy cells^{12,13}, but Vogler *et al.* describe the first that are made up of a protein that can also aggregate in disease. The race is on to search for more of these kinds of functional granule in other cell types. The idea that amyloid-like structures might have beneficial roles, rather than simply being associated with disease, represents a change in our understanding of these protein aggregates. Myo-granules provide a unique opportunity to unravel the differences between a safe and a dangerous aggregate. ■

Lindsay A. Becker and Aaron D. Gitler are in the Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. **L.A.B.** is also in the Stanford Neurosciences Graduate Program,

Stanford University School of Medicine.
e-mail: agitler@stanford.edu

- Forman, M. S., Trojanowski, J. Q. & Lee, V. M.-Y. *Nature Med.* **10**, 1055–1063 (2004).
- Vogler, T. O. *et al.* *Nature* **563**, 508–513 (2018).
- Neumann, M. *et al.* *Science* **314**, 130–133 (2006).
- Ling, S.-C., Polymenidou, M. & Cleveland, D. W. *Neuron* **79**, 416–438 (2013).
- Custer, S. K., Neumann, M., Lu, H., Wright, A. C. & Taylor, J. P. *Hum. Mol. Genet.* **19**, 1741–1755 (2010).
- Eisenberg, D. & Jucker, M. *Cell* **148**, 1188–1203 (2012).
- Nalbandian, A. *et al.* *J. Mol. Neurosci.* **45**, 522–531 (2011).
- Dadon-Nachum, M., Melamed, E. & Offen, D. J. *Mol. Neurosci.* **43**, 470–477 (2011).
- Robinson, J. L. *et al.* *Acta Neuropathol.* **125**, 121–131 (2013).
- Lacorte, E. *et al.* *Neurosci. Biobehav. Rev.* **66**, 61–79 (2016).
- Chiò, A., Benzi, G., Dossena, M., Mutani, R. & Mora, G. *Brain* **128**, 472–476 (2005).
- Boke, E. *et al.* *Cell* **166**, 637–650 (2016).
- Maji, S. K. *et al.* *Science* **325**, 328–332 (2009).

This article was published online on 31 October 2018.

THERAPEUTIC RESISTANCE

A new road to cancer–drug resistance

The discovery of a mechanism that leads to cancer–therapy resistance highlights the many ways that tumour cells can adapt to survive — and reveals the limitations of categorizing patients by their gene mutations. [SEE ARTICLE P.522](#)

KATHARINA SCHLACHER

The development of resistance to cancer therapy is a major predictor of patient mortality. Therefore, understanding resistance mechanisms is key to improving therapeutic outcomes. On page 522, He *et al.*¹ report their discovery of a resistance mechanism in ovarian-cancer cells that contain a mutant version of the *BRCA1* gene.

Mutations in *BRCA1* and *BRCA2* genes can cause breast and ovarian cancer by inactivating either of two major biological pathways that ensure genome stability. One of the pathways repairs DNA double-strand breaks through a process called homologous recombination² (HR). The other process is called fork protection^{3,4}, and safeguards newly synthesized DNA at structures called stalled forks that arise during DNA replication.

In HR repair, an essential bottleneck step is the processing (resection) of double-strand breaks by nuclease enzymes to produce single-stranded (ss) DNA. *BRCA1* acts as a key regulator protein that coordinates the recruitment of the nucleases, which include the MRE11–RAD50–NBS1 protein complex. *BRCA1* also has a second role in HR repair: it recruits *BRCA2*, which in turn loads the

RAD51 protein onto the ssDNA. RAD51 then assists in the binding of the ssDNA to a complementary strand that serves as a template for error-free repair.

Cancer cells that have certain *BRCA1* or *BRCA2* mutations cannot repair double-strand breaks caused by anticancer drugs currently used in the clinic, and so die when treated. Such drugs include cisplatin and PARP inhibitors (PARPi, drugs that specifically target *BRCA*-mutant tumours by taking advantage of their break-repair defects⁵). However, cancer cells can acquire strategies to circumvent the drugs' actions, causing resistance and limiting the use of these initially effective drugs.

In their rigorous study, He *et al.* used a gene-editing screening method⁶ to identify resistance mechanisms in *BRCA1*-mutant ovarian-cancer cells. A known resistance pathway in both *BRCA1*-mutant and *BRCA2*-mutant cells is the restoration of *BRCA* function by re-mutating the original *BRCA* mutation (see ref. 7, for example; Fig. 1a). A second mechanism is drug avoidance, in which a membrane protein pumps the drug out of the cell or reduces its uptake⁸. He and colleagues' screen correctly identified a membrane protein implicated in the uptake of cisplatin by tumour cells as a contributor to resistance, verifying

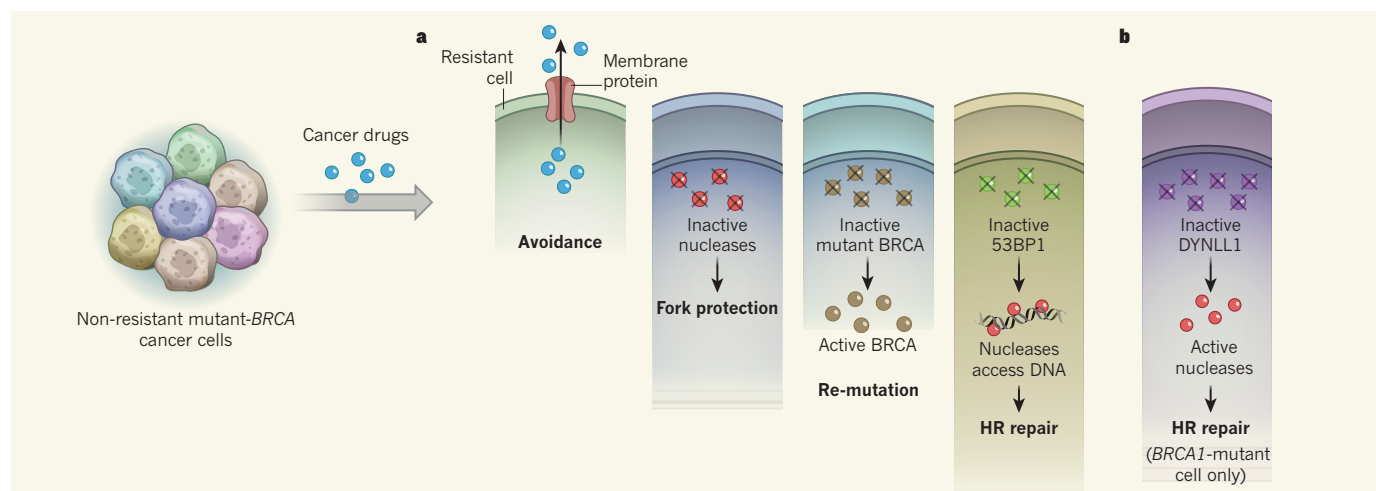


Figure 1 | Mechanisms of drug resistance in cancer cells that contain BRCA mutations. Many cancer cells have mutations in the *BRCA1* or *BRCA2* genes. These mutations inactivate a DNA-repair pathway that involves a process called homologous recombination (HR), or a process known as fork protection that is involved in DNA replication. **a**, *BRCA*-mutated cancer cells have developed many different paths to resist being killed by cancer drugs, including: drug avoidance by pumping drugs out of the cell through a membrane protein; restoration of fork

protection by inactivating nuclease enzymes; re-mutating the original *BRCA* mutation to restore the functions of the *BRCA* proteins; and restoration of HR repair, for example by inactivating the 53BP1 protein in *BRCA1*-mutant cells to allow nucleases to access DNA. Here, the resistant cells are derived from non-resistant cells of the same colour. **b**, He *et al.*¹ report that in *BRCA1*-mutant cells, but not in *BRCA2*-mutant cancer cells, inactivation of the DYNLL1 protein activates nucleases and thus restores HR repair.

the suitability of the authors' approach.

Importantly, one of the top gene 'hits' identified by the screen as causing resistance to both cisplatin and PARPi was *DYNLL1*. The DYNLL1 protein acts in many cellular processes⁹, including intracellular transport and motility, and also inhibits the enzyme nitric oxide synthase (which produces the cell-signalling molecule nitric oxide), but had not been previously implicated in cancer-drug resistance. Deciphering how its inactivation leads to resistance therefore seemed a daunting task.

The authors robustly established that DYNLL1 acts as a negative protein regulator of DNA-end-processing nucleases — it directly interacts with MRE11 and thereby keeps its nuclease activity in check. Inactivation of DYNLL1, therefore, unleashes the nuclease activity of MRE11, even when BRCA1 is not there to help guide it to breaks, and so restores the first of the two HR-repair functions normally carried out by BRCA1 (Fig. 1b).

Conceptually, drug resistance associated with DYNLL1 inactivation is analogous to that caused by inactivation of 53BP1 — another protein that inhibits DNA-end resection, in this case by blocking the access of nucleases to DNA. Inactivation of 53BP1 has been reported to restore resection and therefore resistance in *BRCA1*-mutant cells, but not in *BRCA2*-mutant cells^{10,11}. Moreover, DYNLL1 is known¹² to interact with 53BP1. Yet, unexpectedly, He *et al.* show that resistance associated with DYNLL1 inactivation does not occur through loss of the 53BP1–DYNLL1 interaction.

He and colleagues' study highlights the intricacy of distinct gene functions in cancer-drug resistance, and the importance of

defining biological mechanisms and activities to predict whether tumour cells will be killed. In this case, DYNLL1 inactivation reactivates resection, which is ablated in *BRCA1*-mutant cells, but not in *BRCA2*-mutant tumour cells. DYNLL1 inactivation, therefore, results in resistance in *BRCA1*-mutant cells, but not in *BRCA2*-mutant cells.

Another resistance mechanism that separates BRCA1 and BRCA2 functions has been reported in *BRCA2*-mutant ovarian-cancer cells¹³, where inhibition of the EZH2 enzyme reduces the recruitment of the MUS81 nuclease to replication forks. Notably, this does not restore HR repair, but instead restores fork protection and the survival of *BRCA2*-mutant cells, and not of *BRCA1*-mutant cells. By contrast, EZH2 inhibition increases the sensitivity of *BRCA1*-mutant breast cancers to PARPi (ref. 14). Yet restoration of fork protection by inhibition of MRE11 results in resistance to PARPi and cisplatin in both *BRCA1*- and *BRCA2*-mutant cells, even when HR repair remains defective¹⁵. Thus, it is tempting to suggest that tumour-cell heterogeneity¹⁶ — the existence of tumour-cell subtypes that stem from the same mutation, but which have mutated further to form distinct subpopulations — could partly arise as a result of cells making individual 'decisions' about which survival mechanisms to use, including HR repair, fork protection or both.

Although BRCA1 promotes resection during HR repair, it can also prevent the degradation of newly synthesized DNA by nucleases during fork protection, through an unknown mechanism. These dual, and seemingly opposing, modes of action raise the possibility that restoration of HR repair could promote tumour-cell survival even when fork protection is

dysfunctional. Similarly, DYNLL1 inactivation might in fact cause defects in fork protection by promoting excessive MRE11 activity at replication forks, irrespective of *BRCA* mutations. Future studies will be crucial to dissect the apparently opposing nuclease processes at breaks and forks, and their effects on tumour-cell survival, as a possible nexus point for therapeutic intervention.

The emergence of diverse mechanisms for cancer-drug resistance demonstrates that cancer cells respond distinctively to individual defects of molecular function — rather than to an overall genetic defect — to rebalance the cellular homeostasis that ensures their survival. He *et al.* identified DYNLL1 inactivation as a resistance mechanism to cisplatin and PARPi; although both drugs cause double-strand breaks, their mode of action differs. In addition, other commonly used anticancer drugs, including gemcitabine, 5-fluorouracil and hydroxyurea, mainly disrupt replication reactions. Researchers, therefore, should expect to identify many different ways in which resistance can develop. These roads to resistance might require the restoration of replication processes, repair processes or both. The sequential use of different cancer therapies when an initial treatment is not successful is routine practice, but could lead to the development of multiple resistance mechanisms, and ultimately to resistance to any of the therapies.

People with cancer who have *BRCA1* and *BRCA2* mutations are currently grouped together in many genome studies and when considering treatment options, despite increased understanding of the molecular and genetic distinctions. He and co-workers' study suggests that molecular function, rather than genotype function — in this case, the specific

role of BRCA1 in resection, rather than its overall role in HR repair or in fork protection — dictates the cellular outcomes. More broadly, these results suggest that personalized-medicine strategies should be considered that take into account molecular functions in individuals, rather than categorizing people solely by genotype. ■

Katharina Schlacher is in the Department of Cancer Biology, University of Texas

MD Anderson Cancer Center, Houston, Texas 77058, USA.

e-mail: kschlacher@mdanderson.org

1. He, Y. J. *et al. Nature* **563**, 522–526 (2018).
2. Moynahan, M. E. & Jasin, M. *Nature Rev. Mol. Cell Biol.* **11**, 196–207 (2010).
3. Schlacher, K. *et al. Cell* **145**, 529–542 (2011).
4. Schlacher, K., Wu, H. & Jasin, M. *Cancer Cell* **22**, 106–116 (2012).
5. Bryant, H. E. *et al. Nature* **434**, 913–917 (2005).
6. Shalem, O. *et al. Science* **343**, 84–87 (2014).
7. Sakai, W. *et al. Cancer Res.* **69**, 6381–6386 (2009).
8. Goldstein, L. J. *et al. J. Natl Cancer Inst.*

81, 116–124 (1989).

9. Barbar, E. *Biochemistry* **47**, 503–508 (2008).
10. Bunting, S. F. *et al. Cell* **141**, 243–254 (2010).
11. Bouwman, P. *et al. Nature Struct. Mol. Biol.* **17**, 688–695 (2010).
12. Lo, K. W. *et al. J. Biol. Chem.* **280**, 8172–8179 (2005).
13. Rondinelli, B. *et al. Nature Cell Biol.* **19**, 1371–1378 (2017).
14. Yamaguchi, H. *et al. Oncogene* **37**, 208–217 (2018).
15. Ray Chaudhuri, A. *et al. Nature* **535**, 382–387 (2016).
16. Sottoriva, A. *et al. Nature Genet.* **47**, 209–216 (2015).

This article was published online on 31 October 2018.

OPTOELECTRONICS

Efficiency breakthrough for radical LEDs

A strategy for using organic free radicals to make light-emitting diodes circumvents the constraints that limit the efficiency with which other organic LEDs convert electric current into light. [SEE LETTER P.536](#)

TETSURO KUSAMOTO & HIROSHI NISHIHARA

Light-emitting devices made from organic materials have the potential to be thin, flexible and lightweight, and might therefore be used in a variety of applications — including foldable display screens, ‘smart’ wallpaper incorporating digital devices, and windows that could be converted into illuminated panels at the flick of a switch. On page 536, Ai *et al.*¹ report the development of organic light-emitting diodes (OLEDs) that use free radicals as the emitter and convert

electrons into light with high efficiency. The efficiencies of other types of OLED are generally limited by quantum-mechanical effects, but radical-based OLEDs (ROLEDs) don’t have this constraint, owing to the electronic state of the radicals. The authors’ ROLEDs have the highest emission efficiency obtained so far among LEDs that emit light in the deep-red and infrared regions of the electromagnetic spectrum.

Several types of LED are being actively developed because they are expected to produce displays that have higher brightness,

colour purity, contrast and resolution than conventional lighting devices, while using less energy. OLEDs, in particular, have become familiar in the past decade, because they are used in the displays of mobile phones and televisions. Such displays perform better in several respects (such as contrast and colour reproducibility) than do liquid-crystal displays, which are currently used in many electronic devices.

OLEDs were first reported² in 1987, and typically have a multilayered structure: a layer of material that contains light-emitting molecules is sandwiched between layers that transport electrons and holes (positively charged quasiparticles formed by the absence of electrons in atomic lattices), which, in turn, are sandwiched by electrodes as the outermost layers (Fig. 1). Additional layers that enable efficient injection of holes and electrons from the electrodes into the transport layers are also sometimes used. When an electric field is applied between the two electrodes, holes and electrons are injected and merge (recombine) on emitter molecules in the light-emitting layer to generate photons. The structure of the emitter molecule determines the colour of the emission.

One problem that still needs to be overcome for OLEDs is their low efficiency, which is quantified by the external quantum efficiency (EQE) — the ratio of the number of photons that leave the device to the number of electrons injected into it on the application of an electric field. The EQE is, in turn, proportional to two factors: the internal quantum efficiency (IQE), which is the efficiency with which photons are generated in the light-emitting layer from injected electrons; and the light outcoupling efficiency, which is the ratio of the number of photons that exit the device to the number generated within it. The value of the outcoupling efficiency is typically 20–30% (ref. 3). Quantum mechanics dictates that the IQE of conventional OLEDs based on fluorescent molecules is limited to 25% (ref. 4). The remaining 75% of efficiency is lost through recombination pathways that don’t result in light emission. The EQEs of such OLEDs are therefore 5–6% at best.

Several groundbreaking methods have been established to solve the efficiency problem. For example, the IQE of OLEDs has been increased to nearly 100% by using phosphorescence (rather than fluorescence) as the light-emitting process⁵, or by using a heat-activated

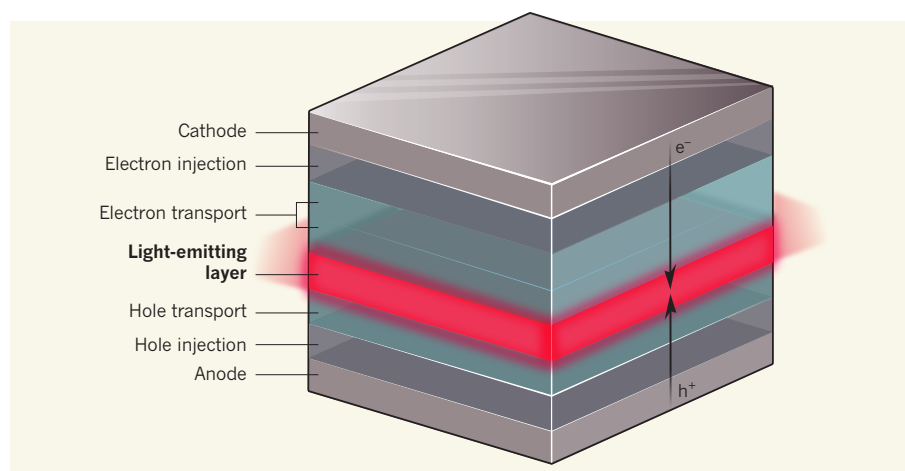


Figure 1 | An efficient radical-based organic light-emitting diode (ROLED). Ai *et al.*¹ report two organic free radicals that can be used in multilayered light-emitting diodes. Electrons (e^-) and holes (h^+ ; quasiparticles formed by the absence of electrons in an atomic lattice), which are produced by a cathode and an anode, respectively, pass through injection layers and transport layers before merging (recombining) on radical molecules in the light-emitting layer. This recombination produces light in the deep-red and infrared regions of the electromagnetic spectrum. Photons are produced from electrons in the light-emitting layer with almost 100% efficiency. The maximum external quantum efficiency of the device (the ratio of the number of photons that leave the LED to the number of electrons injected into it) is 27%, the highest such efficiency of any LED that emits deep-red and infrared light.

light-emitting mechanism known as delayed fluorescence⁶. These strategies overcome the problem for devices based on conventional fluorescent molecules, but Ai *et al.* now report an innovative alternative method: they use organic radical molecules that exploit a different light-emitting mechanism, thereby enabling an IQE of almost 100%.

So what are organic radicals? Most organic molecules have an even number of electrons, in which each electron pairs up with another one, forming what is known as a closed-shell state. Organic radicals, however, have an odd number of electrons, and have one or more unpaired electrons in 'open-shell' states. Such radicals are highly reactive and therefore chemically unstable, and are typically generated transiently during chemical reactions. But the reactivity of radicals can be suppressed by modifying their molecular structures, and some are stable enough to be handled under air at room temperature.

In the context of light emission, it has long been thought that almost all stable radicals are non-emissive and inhibit emission from other sources. Nevertheless, stable light-emitting radicals have been available^{7,8} since 2006, raising the possibility that they could be used in lighting materials and devices. Importantly, it was proposed⁹ that ROLEDs would have high IQEs because, owing to the radicals' open-shell electronic states, they don't exhibit the energy-loss pathways that cause problems in conventional OLEDs.

The first ROLED was reported¹⁰ in 2015 by researchers from one of the groups that contributed to the current paper, and it had an EQE of 2.4%. A year later, the same group showed experimentally³ that it should be possible for ROLEDs to achieve an IQE of 100% — a milestone in the history of this LED class. Ai *et al.* now report another key step in the evolution of ROLEDs: they have developed two stable radicals that emit brightly in the deep-red and infrared regions of the spectrum, and they use them in devices that not only achieve almost 100% IQE, but also have an excellent EQE of 27%. This is the highest EQE among all LEDs that emit similar colours, and is largely a consequence of the efficiency with which electrons are converted into light on the radicals.

The high efficiency of Ai and colleagues' device is impressive, but ROLEDs in general currently emit light in only a limited range of colours. This is because just a small number of stable light-emitting radicals have been reported, and only those that have a particular type of chemical structure (known as an electron-donating group) deliver high EQEs when used in ROLEDs. Moreover, the electronic characteristics of light-emitting radicals suggest that these molecules will not be good at emitting blue (high-energy) light. A crucial next step will be to establish molecular design principles that enable organic radicals to be tuned to produce a wide range of colours — Ai and co-workers' radicals are not the first to

emit deep-red and infrared light, and so have not extended the colour range.

Nonetheless, Ai and co-workers have demonstrated an innovative method for increasing the EQE of OLEDs, which could not have been achieved through simple developments of conventional fluorescent OLEDs. The authors' method also increases the number of radicals that can be used in ROLEDs. Given that they were discovered only a few years ago, there is probably plenty of potential for even further improvement — a challenge that offers great opportunities for materials scientists. In this field, radical progress truly promises a bright future. ■

Tetsuro Kusamoto and Hiroshi Nishihara
are in the Department of Chemistry, School

BIOPHYSICS

Membranes stick to one dimension

A nanometre-scale mechanism has been proposed to explain how bacteria improve their grip on human cells. The findings have implications for drug discovery, and might inspire biomimetic applications such as adhesives.

JOHN R. DUTCHER

Biological membranes serve as the barrier between cells and their surrounding environment, and regulate the transfer of ions and small molecules into and out of cells. Because of their central role in proper cellular operation, membranes are a target for many disease-causing microorganisms¹. Writing in *Nature Communications*, Charles-Orszag *et al.*² propose a previously unknown mechanism by which one such pathogenic bacterium, *Neisseria meningitidis* (also known as meningococcus), rearranges the outer plasma membrane of host cells to improve its adhesion to the cells. Achieving improved cell adhesion is a key step in host infection, which in humans can lead to septic shock and meningitis³.

A key challenge for *N. meningitidis* is how to stick to and colonize the endothelial cells that line blood vessels without being swept away by flowing blood. The interaction between the bacterial and endothelial-cell surfaces is not strong enough to withstand the forces exerted by blood flow⁴, and so *N. meningitidis* uses extremely thin (6-nanometre-diameter) protein filaments called type IV pili (T4P) to increase its grip on the cell membrane. T4P can be extended and retracted through the cell wall in a variety of bacteria, and have crucial roles in the microbes' life cycle, allowing them to stick to and move across

of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
e-mail: kusamoto@chem.s.u-tokyo.ac.jp

1. Ai, X. *et al.* *Nature* **563**, 536–540 (2018).
2. Tang, C. W. & VanSlyke, S. A. *Appl. Phys. Lett.* **51**, 913–915 (1987).
3. Obolda, A., Ai, X., Zhang, M. & Li, F. *ACS Appl. Mater. Interfaces* **8**, 35472–35478 (2016).
4. Rothberg, L. J. & Lovinger, A. J. *J. Mater. Res.* **11**, 3174–3187 (1996).
5. Baldo, M. A. *et al.* *Nature* **395**, 151–154 (1998).
6. Uoyama, H., Goushi, K., Shizu, K., Nomura, H. & Adachi, C. *Nature* **492**, 234–238 (2012).
7. Gamaro, V. *et al.* *Tetrahedron Lett.* **47**, 2305–2309 (2006).
8. Heckmann, A. *et al.* *J. Phys. Chem. C* **113**, 20958–20966 (2009).
9. Hattori, Y., Kusamoto, T. & Nishihara, H. *Angew. Chem. Int. Edn* **53**, 11845–11848 (2014).
10. Peng, Q., Obolda, A., Zhang, M. & Li, F. *Angew. Chem. Int. Edn* **54**, 7091–7095 (2015).

surfaces and to infect or damage other cells⁵.

The interaction between *N. meningitidis* cells and endothelial cells results in the formation of protrusions on the endothelial-cell membrane⁴, and it has been shown that proteins in T4P are essential for protrusion formation⁶, and that they interact with specific receptors in the endothelial cells⁷. However, the molecular mechanism underlying the interaction of T4P with host cells was not understood. Charles-Orszag and co-workers now shed light on this mechanism by combining *in vivo* and *in vitro* studies with a simple theoretical model.

The theoretical model is one of the strengths of the new study, and describes a previously unknown mechanism for wetting (the spreading of a deformable substance such as a liquid on the surface of another substance⁸). Wetting is key to many aspects of everyday life, from the spreading of ink on paper to the beading of water droplets on spider webs or freshly waxed cars, and it typically occurs in two dimensions. However, in the case of a very thin fibre in contact with a soft membrane, the membrane cannot wrap around (wet) the fibre, because too much energy is required to accommodate the large curvature around the fibre's cross-section.

Charles-Orszag *et al.* use their model to show that it can be energetically more favourable for a narrow tube to be drawn out from the membrane, along the fibre, than wrapped around it (Fig. 1). This mechanism

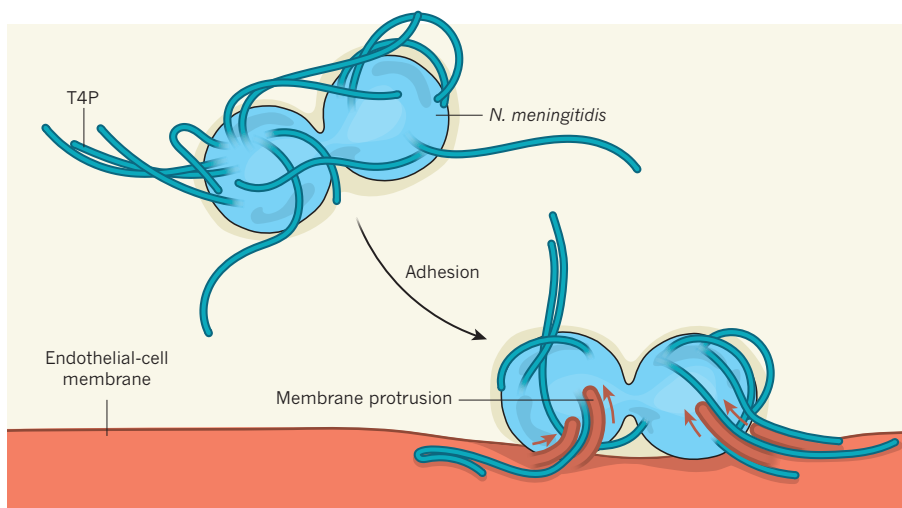


Figure 1 | Enhancing adhesion between a bacterium and an endothelial cell. The bacterium *Neisseria meningitidis* attaches itself to the endothelial cells that line blood vessels in host organisms. The bacterium uses fibres known as type IV pili (T4P) to induce the formation of protrusions from endothelial-cell membranes. These protrusions strengthen the bacterium's hold on the membrane, helping it to colonize cells without being swept away by the surrounding blood flow. Charles-Orszag *et al.*² propose that the adhesion of T4P to the membrane drives a process called one-dimensional wetting, in which the protrusions are drawn along the T4P fibres (red arrows). (Adapted from Fig. 5 of ref. 2.)

for forming membrane protrusions, which the authors call one-dimensional wetting, is driven by adhesion between the membrane and the fibre. The membrane protrusions could help to anchor a bacterium to a host cell as its T4P extend and retract, without breaking the adhesive interactions between the T4P and the membrane — thus maintaining the dynamic nature of the fibres.

Because the remodelling of endothelial-cell membranes by *N. meningitidis* had previously been observed only for cultured cells, the researchers studied blood vessels in human skin grafted onto mice to confirm that remodelling also occurs *in vivo*. They then complemented those experiments with *in vitro* studies to explore the mechanism involved. Unfortunately, the *in vitro* experiments did not examine the interaction of isolated T4P with model membranes, because this would have required the appropriate receptor proteins to be introduced into the membranes. Instead, Charles-Orszag *et al.* studied two model systems: artificial cells (known as giant unilamellar vesicles) interacting with filaments of a protein called actin through adhesion between the filaments and molecules attached to the cells; and endothelial-cell membranes interacting with mimics of the fibres found in the extracellular matrix around cells.

The authors show that 1D wetting does indeed occur in these systems, and that it can be understood quantitatively using their model. Their *in vitro* observations highlight the essential feature of this phenomenon: the presence of adhesion between a deformable membrane and a nanoscale fibre. Their observations also suggest that 1D wetting could occur more generally for physiologically important interactions of human cells with

other biological nanofibres, and that it could have a major role in cell migration.

Further work is needed to understand 1D wetting in more detail. Systematic studies in which the fibre radius, strength of the adhesive interaction and surface tension of the membrane are varied would improve our understanding. In addition, further developments in microscopy will lead to better

visualization of the structure and dynamics of the protrusions involved in 1D wetting.

Charles-Orszag and co-workers' results reveal opportunities for biomimetic strategies for wetting synthetic nanofibres and for producing strong adhesives, and new ways of moving nanoscale objects. Their findings also imply that reducing or disabling the 1D wetting of *N. meningitidis* T4P would limit the bacterium's ability to colonize and infect host cells, opening up a potential avenue for drug discovery. More generally, 1D wetting might enable cell function and health to be manipulated through interactions of cells with nanofibres to which biologically active molecules have been attached. ■

John R. Dutcher is in the Department of Physics and the Biophysics Interdepartmental Group, University of Guelph, Ontario N1G 2W1, Canada. e-mail: dutcher@uoguelph.ca

1. Alberts, B. *et al.* *Molecular Biology of the Cell* 5th edn (Garland Science, 2008).
2. Charles-Orszag, A. *et al.* *Nature Commun.* **9**, 4450 (2018).
3. Coureuil, M. *et al.* *Virulence* **3**, 164–172 (2012).
4. Mikaty, G. *et al.* *PLoS Pathog.* **5**, e1000314 (2009).
5. Burrows, L. L. *Mol. Microbiol.* **57**, 878–888 (2005).
6. Melican, K., Michea Veloso, P., Martin, T., Bruneval, P. & Dumenil, G. *PLoS Pathog.* **9**, e1003139 (2013).
7. Maissa, N. *et al.* *Nature Commun.* **8**, 15764 (2017).
8. De Gennes, P.-G., Brochard-Wyart, F. & Quere, D. *Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves* (Springer, 2004).

This article was published online on 5 November 2018.

GENETICS

A genomic approach to mosquito control

A high-quality genome sequence for the mosquito *Aedes aegypti* has now been assembled. The sequence will enable researchers to identify genes that could be targeted to keep mosquito populations at bay. [SEE ARTICLE P.501](#)

SUSAN E. CELNIKER

Every year, millions of people are bitten by the mosquito *Aedes aegypti*. Thousands die as a result of infection by the viruses the mosquito carries¹, which can cause diseases such as yellow fever, dengue fever and Zika. Current mosquito-suppression methods typically involve pesticides. However, mosquitoes quickly develop resistance to these chemicals², and pesticides can accumulate in the food chain, with adverse effects on beneficial insects, other wildlife and humans. New control methods are therefore needed. On page 501, Matthews *et al.*³ describe a high-quality genome sequence for *A. aegypti* (Fig. 1).

This exemplary work could be a major step towards addressing our current inability to manage expanding mosquito populations.

Arguably the most promising alternatives to pesticide-based mosquito control are targeted molecular strategies based on genetics. The first requirement for the success of such strategies is high-quality sequencing of the mosquito genome. This would enable researchers to identify gene targets that could be manipulated to achieve a range of effects: to disrupt the mosquito's host-targeting systems; to make sterile males; to convert females into harmless males; or to render the insect incapable of harbouring viruses.

The repetitive nature of the 1.3-gigabase-long

A. aegypti genome has severely hampered efforts to generate a high-quality sequence. Previous attempts^{4,5} resulted in patchy genomes that were assembled using short sequence reads. To overcome these challenges, Matthews *et al.* used next-generation sequencing to generate 166 Gb of long sequence reads with an average length of 17 kilobases. The authors used sophisticated mapping and gap-filling techniques to determine the positions of 94% of their sequence reads on the mosquito's three chromosomes, successfully assembling 1.28 Gb of the genome. The assembly has many fewer gaps than previous assemblies, and is a 100-fold improvement in terms of its N50 — a statistical measure based on the median assembled DNA-sequence length.

With this assembly in hand, Matthews and colleagues were able to improve our knowledge of the sequences of thousands of genes, and to discover new members of existing gene families. For example, the researchers identified more than 300 genes that encode ligand-gated ion channels, which allow ions to pass through membranes. These genes fall into three classes of receptor: odorant, gustatory and ionotropic. Together, they sense a wide range of chemicals, including carbon dioxide and chemicals that emanate from humans. Matthews *et al.* identified 54 previously unknown genes encoding ionotropic receptors — almost doubling the number known before. These genes are ideal candidates to target for disruption, because they confer the mosquito's ability to detect odours that indicate the presence of a host.

Of note, the authors identified 14 members of the best-studied subgroup of ionotropic receptors, nicotinic acetylcholine receptors, which act in the insect nervous system⁶. These receptors are the targets of insecticides called neonicotinoids, which have gained much attention owing to their adverse effects on beneficial insects such as bees. Knowing the sequences of the genes that encode these receptors should enable researchers to design insecticides that specifically target mosquitoes, sparing beneficial species.

Gene duplication is one mechanism by which insects can develop resistance to pesticides. Matthews *et al.* used their assembly to resolve a complicated gene-repeat region involved in one such resistance event. The region contains a cluster of three *Glutathione S-transferase* (*GST*) genes, which the authors found had been duplicated four times. These genes are important for metabolizing toxins, with one gene, *GSTe2*, capable of metabolizing the insecticide DDT. Increased expression of *GSTe2* has been associated with DDT resistance in a laboratory-colonized *A. aegypti* strain⁷, supporting the idea that the gene duplication identified by the authors is involved in pesticide resistance. These data provide a proof of principle that the new genome will be an invaluable resource for researchers looking to analyse any gene family implicated in pesticide resistance.

Sex determination in *A. aegypti* is controlled



Figure 1 | The mosquito *Aedes aegypti*. Matthews *et al.*³ describe a high-quality genome sequence for this mosquito species.

by a sex-specific region called the M locus that is located on chromosome 1 in males only. It was known that the region contained the male-specific genes *myo-sex* and *Nix*, but they were absent from previous genome assemblies. This gap has been filled in the new genome. The authors estimate the M locus to be 1.5 megabases long (0.1% of chromosome 1), and show that it contains a much more repetitive sequence than does the rest of the genome — 73.7% compared with 11.7% genome-wide. The high repeat density is similar to that found in the Y chromosome of other animals⁸.

Apart from the M locus, the sequence of chromosome 1 is very similar in males and females. This type of chromosome structure is known as homomorphic. Matthews and colleagues' genome will provide researchers with the opportunity to examine how the homomorphic sex chromosomes of *A. aegypti* are maintained, rather than evolving into heteromorphic chromosomes that are broadly different between the sexes — a better-understood phenomenon that is exemplified by the human X and Y chromosomes.

Finally, the authors used genetic-mapping techniques to identify regions of the genome that are associated both with the ability of mosquitoes to act as vectors for dengue virus and with resistance to the pesticide deltamethrin. The latter analysis highlighted candidate genes not previously known to be involved in pesticide resistance.

Even though Matthews and co-workers' genome is a radical improvement on previous assemblies, important genes might still be missing, because there are a few thousand gaps in the main chromosomes, and large gaps spanning specialized structures called centromeres, to which proteins bind during cell division. Nonetheless, the authors' sophisticated

genome-sequencing strategy should act as a template for future efforts to assemble complex genomes. The genome and the gene sets themselves are publicly available for others to use (see go.nature.com/2dc6kxp), and, thanks to genome-editing technologies such as CRISPR-Cas9, researchers will easily be able to explore the effects of disrupting each gene identified as a candidate for targeting.

The use of tools rooted in genomic analysis and manipulation is a key step towards a pesticide-free world. Matthews and colleagues' work makes a major contribution to this goal. ■

Susan E. Celniker is in the Department of BioEngineering & BioMedical Sciences, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
e-mail: secelniker@lbl.gov

1. Byard, R. W. *Am. J. Forens. Med. Pathol.* **37**, 74–78 (2016).
2. Moyes, C. L. *et al. PLoS Negl. Trop. Dis.* **11**, e0005625 (2017).
3. Matthews, B. J. *et al. Nature* **563**, 501–507 (2018).
4. Nene, V. *et al. Science* **316**, 1718–1723 (2007).
5. Dudchenko, O. *et al. Science* **356**, 92–95 (2017).
6. Rudloff, E. *Exp. Cell Res.* **111**, 185–190 (1978).
7. Lumjuan, N., McCarroll, L., Prapanthadara, L., Hemingway, J. & Ranson, H. *Insect Biochem. Mol. Biol.* **35**, 861–871 (2005).
8. Smith, K. D., Young, K. E., Talbot, C. C. Jr & Schmeckpeper, B. J. *Development* **101**, 77–92 (1987).

This article was published online on 14 November 2018.

CORRECTION

The News and Views article 'Beating the quantum limits (cont'd)' (*Nature* **331**, 559; 1988) gave the wrong citation for Masanao Ozawa's paper. It should have referred to M. Ozawa *Phys. Rev. Lett.* **60**, 385 (1988).

Structural superlubricity and ultralow friction across the length scales

Oded Hod^{1*}, Ernst Meyer², Quanshui Zheng^{3*} & Michael Urbakh¹

Structural superlubricity, a state of ultralow friction and wear between crystalline surfaces, is a fundamental phenomenon in modern tribology that defines a new approach to lubrication. Early measurements involved nanometre-scale contacts between layered materials, but recent experimental advances have extended its applicability to the micrometre scale. This is an important step towards practical utilization of structural superlubricity in future technological applications, such as durable nano- and micro-electromechanical devices, hard drives, mobile frictionless connectors, and mechanical bearings operating under extreme conditions. Here we provide an overview of the field, including its birth and main achievements, the current state of the art and the challenges to fulfilling its potential.

Friction is one of the oldest phenomena examined and used by humankind¹. It has diverse implications in many scientific and technological fields, ranging from physics and chemistry to biology and engineering^{2–5}. In the macroscopic world, friction is an inherent phenomenon in the operation of any mechanical system. Whereas in some cases it is essential for the proper function of the device, friction is often responsible for considerable energy loss and wear⁶. In fact, it has been estimated that about one-third of the energy supplied by fossil fuel in automotive vehicles is consumed in overcoming various forms of frictional dissipation⁷. Friction-induced wear becomes a severe problem when reducing the system size to the nanoscale. Here, owing to the intrinsically high surface-to-volume ratio, even the slightest surface wear may hinder device operation and reduce its durability. One may naively suggest traditional lubrication approaches as a remedy for this problem. However, standard liquid-phase lubricants—for example, organic oils—have been shown to either become highly viscous when confined to nanoscale constrictions⁸ or completely evacuate the junction under external pressure leaving behind a bare frictional interface⁹. As the world strives to miniaturize electronic and mechanical technologies towards ever smaller length scales, new approaches are therefore required to decrease or even eliminate friction and wear at reduced dimensions.

The natural world suggests many alternative strategies for friction reduction. The most common example is lubricated friction, where surface fluid molecules adsorbed at the sliding interface serve as a tribological buffer layer. For instance, polymer brushes and water solvation shells have been suggested to provide the remarkable durability of skeletal joints operating under extreme loads^{4,10–12}. In this respect, the superlubric properties of such brush architectures have been investigated down to the single molecule level¹³. Recently, ionic liquids have emerged as alternative fluid lubricants allowing electro-tunable ultralow friction to be achieved in non-aqueous environments^{15,16}. Solid lubrication constitutes a different approach to the reduction of friction, and relies on the introduction of micrometre-scale or nanoscale particles into the contact region¹⁷. These particles can serve as miniature bearings as well as a source of lubricating flakes via successive layer exfoliation or complete collapse of onion-type structures¹⁷.

The schemes described above follow the standard paradigm that friction reduction requires the introduction of external lubricants into the sliding junction. Notably, ultralow friction can be achieved in the

complete absence of such lubricants. One such scheme, first studied in the early 1970s, involved the use of layered materials, such as graphite, for ultralow friction substrates^{18,19}. This approach enabled the reduction of kinetic friction coefficients down to 5×10^{-3} at nano- and micro-scale contacts under relatively low loads. More recently, amorphous diamond-like carbon appeared as a promising coating for achieving super-low friction and wear at even larger-scale junctions²⁰. Such films are already used in many industrial applications, including razor blades, magnetic hard drives, engine parts, mechanical face seals, scratch-resistant glasses, invasive and implantable medical devices, as well as micro-electromechanical systems²¹. An alternative approach to the reduction of dry friction and wear involves mechanical modulation of the normal and lateral forces applied to the interface, resulting in the elimination of stick-slip motion and hence decrease in energy dissipation^{22–24}.

In this Perspective, we focus on a different, inherent, type of lubricant-free friction reduction scheme that appears in incommensurate crystalline contacts. This mechanism, often termed structural superlubricity²⁵, is one of the most interesting concepts in modern tribology, and holds promise for the achievement of even lower friction coefficients²⁶. It relies on the lattice misfit between two flat and rigid crystalline surfaces leading to effective cancellation of the lateral forces during sliding motion²⁷. The major advantage of structural superlubricity is the ability to circumvent the need for external additives or mechanical manipulation by using chemically clean and stiff surfaces that preserve their crystal lattice structure under shear stress, thus maintaining their incommensurate configuration. This provides intrinsic lubrication that can be adjusted, by the nature of the contacting materials and the junction geometry, to be robust even under extreme conditions, such as high pressures, high and low temperatures, as well as vacuum.

The realization and characterization of structural superlubricity at microscale contacts recently became feasible with advances in the fabrication of large-scale pristine single-crystal layered materials and the development of supersensitive manipulation devices. This constituted an increase of scale of three orders of magnitudes with respect to nanoscale contact experiments performed only a decade ago. In combination with recent advances in the computational modelling of such junctions, this marks out the field of structural superlubricity research as timely and exciting with great promise for realistic technological applications. Here we provide a general overview of the field, starting from its inception,

¹Department of Physical Chemistry, School of Chemistry, The Raymond and Beverly Sackler Faculty of Exact Sciences, and The Sackler Center for Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv, Israel. ²Department of Physics, University of Basel, Basel, Switzerland. ³Department of Engineering Mechanics, Center for Nano and Micro Mechanics, Applied Mechanics Laboratory, and State Key Laboratory of Tribology, Tsinghua University, Beijing, China. *e-mail: odedhod@tau.ac.il; zhengqs@tsinghua.edu.cn

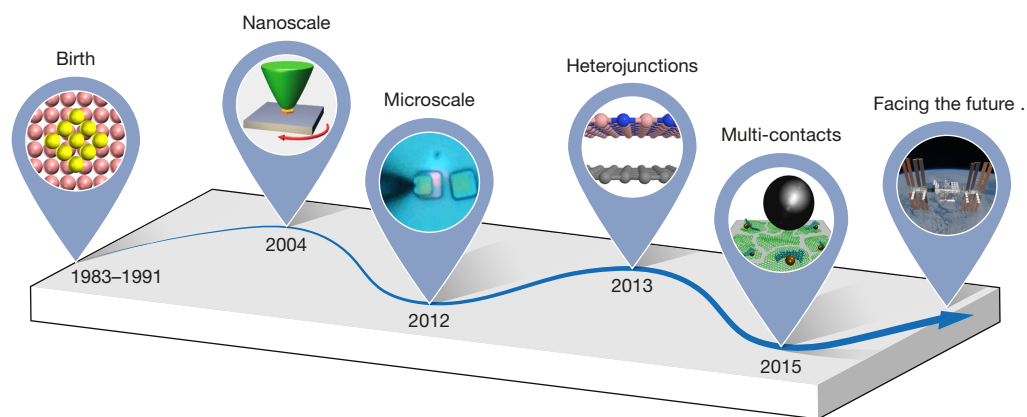


Fig. 1 | Timeline of major milestones in structural superlubricity research. The timeline starts with the first theoretical prediction of vanishing static friction, made in 1983²⁸, and the computational study of ultralow kinetic friction states in 1991²⁹ ('Birth'). This is followed by the pioneering experimental demonstration of nanoscale superlubricity in graphitic contacts in 2004⁴¹, which led to the first observations of microscale superlubricity in 2012⁴² and to the suggestion of heterojunctions⁸⁶ in 2013 and of multi-contact configurations⁷ in 2015 (multi-contacts schematic image adapted with permission from ref. ⁷,

American Association for the Advancement of Science) as possible routes to achieve robust superlubricity at large length scales. This path taken by the scientific community in recent years opens the door to the scaling-up of structural superlubricity towards the macroscale, with substantial technological implications and applications, such as solid lubricants for satellite solar panel motors operating under the extreme conditions encountered in space ('Facing the future ...'; satellite image adapted from https://www.nasa.gov/multimedia/imagegallery/image_feature_1314.html, NASA).

discuss the main achievements and the state of the art, and foresee the challenges that are yet to be overcome towards achieving this goal. We aim to turn the attention of the scientific community to this phenomenon and to trigger new fundamental and applied research for its scaling-up to the macroscale.

The birth of structural superlubricity

The first theoretical prediction of such a state of vanishing static friction in crystalline interfaces was given by Peyrard and Aubry²⁸ for infinite incommensurate contacts in 1983 (see Fig. 1). The term superlubricity was coined by Hirano and co-workers²⁹ almost a decade later (see Fig. 1), referring to the suppression of stick-slip motion via the elimination of a particular energy dissipation channel related to elastic instabilities. Such stick-slip dynamics, commonly associated with the squeaky sound of opening unoled doors (widely used in horror movies), is a major source of energy dissipation, hence its suppression results in considerable reduction of dynamic friction. Nevertheless, in practice, there always exist alternative energy dissipation routes (for instance, the excitation of lattice vibrations induced by variations of long-range interactions) and wear mechanisms resulting in residual friction even in the absence of stick-slip motion. Therefore, unlike other critical phenomena, such as superconductivity and superfluidity, frictional energy dissipation never truly vanishes. In light of this, the criterion for the onset of superlubricity is commonly chosen as the reduction of the friction coefficient (the derivative of the friction force with respect to the normal load) to below 10^{-3} – 10^{-4} .

Insight into the phenomenon of structural superlubricity can be gained by considering the interactions between two surfaces made from plastic foam, each bearing an 'egg-box' pattern of peaks and troughs (see Fig. 2a, b). When the corrugated surfaces of the two foams are put in registry, one can hardly induce lateral sliding because many high barriers have to be crossed simultaneously over the entire interface. Nevertheless, when one foam is slightly laterally rotated with respect to the other, the lattices are taken out of registry. In this case, when one surface slides upon the other some of its peaks are forced to climb uphill while others go downhill. For sufficiently large interfaces these local opposite motions result in effective cancellation of the global friction force. A similar mechanism holds true for micro- and nanoscale interfaces, with the corrugated foam surfaces being replaced by the potential energy landscape of the inter-surface interactions (see Fig. 2c).

Naturally, realistic material interfaces are more complicated than implied by the rigid egg-box foam model. Specifically, the elasticity

of contacting materials may affect superlubric behaviour. Such effects are already appearing in single-particle phenomenological treatments such as the Prandtl–Tomlinson model, where a point mass, dragged by an external support via an elastic spring of stiffness k , slides atop a periodic sinusoidal potential of periodicity a_0 and amplitude V_0 , representing the underlying surface^{30,31}. Here, a transition from stick-slip motion to smooth sliding occurs when the dimensionless parameter $\eta = 4\pi V_0/(ka_0^2)$ exceeds the critical value $\eta = 1$. At this point the mechanical instability resulting from the competition between the driving spring force and the opposing frictional force, exerted by the potential energy landscape, is eliminated. A more realistic description of contacting surfaces requires extension to a multi-particle treatment, such as the Frenkel–Kontorova model³². This introduces intra-surface elasticity that allows the slider atoms to accommodate to the underlying potential, as directly demonstrated by a recent experiment using cold atom chains residing on an optical lattice³³. As a result, above a critical contact size that depends on the ratio between the intra-surface elasticity and interfacial stiffness, locally commensurate regions may form, leading to pinning effects and enhancement of friction^{25,34–36}. Notably, this theoretical prediction was recently verified experimentally³⁷ for antimony particles sliding atop MoS₂. We note that such pinning effects are not limited to the context of tribology but are of rather general nature and well known to the superconductivity community^{25,38}. An important advantage of layered materials is their extremely stiff intra-layer structure and low inter-layer potential energy surface corrugation that may shift the critical length to macroscopic scales.

Experimental evidence of structural superlubricity was reported as early as in 1993, for homogeneous MoS₂ interfaces³⁹. This was further supported by experiments on nanoscale heterogeneous MoS₂/MoO₃ junctions, which exhibited the anisotropic friction characteristic of these systems⁴⁰. A decade later (see Fig. 1), the first detailed experimental exploration of the mechanisms of structural superlubricity in nanoscale graphitic contacts was undertaken, demonstrating controllable and reproducible superlubric motion⁴¹. This triggered extensive experimental investigations that resulted in promising realizations of superlubricity in microscopic graphitic contacts as well as in centimetre-long carbon nanotubes, as detailed below^{42–45}. These impressive recent advances constitute important milestones towards the achievement of macroscale superlubricity, which holds great technological promise for the reduction of friction and wear in actual mechanical devices. Nevertheless, with increasing contact size, factors such as in-plane elasticity and out-of-plane corrugation as well as surface defects and

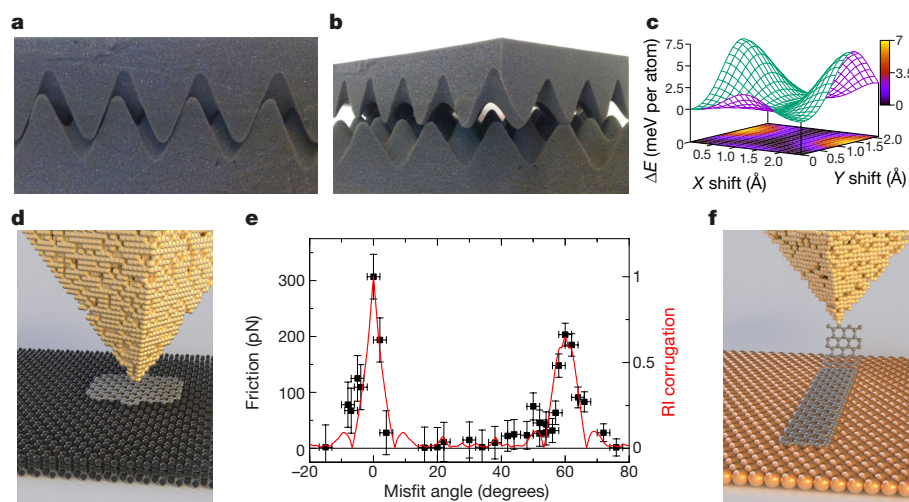


Fig. 2 | Nanoscale structural superlubricity. **a, b**, Egg-box foam models of commensurate **(a)** and incommensurate **(b)** lattices illustrating the origin of finite and vanishing interfacial friction states, respectively (both images reproduced with permission from ref. ⁴⁷, American Physical Society). Each individual foam surface represents the potential energy landscape experienced by a single atom sliding atop the atomic lattice of a rigid two-dimensional layered surface. **c**, The resulting inter-layer sliding energy landscape of a graphene bilayer, obtained by a dedicated inter-layer potential calibrated against advanced density functional theory calculations (image adapted with permission from ref. ⁹⁹, American

Chemical Society). **d, e**, The first experimental demonstration of such a transition from finite to vanishingly small friction was obtained for graphitic flakes sliding atop graphite **(d)**, where friction was shown to reduce by orders of magnitude upon relative rotation of the originally aligned contacting surfaces⁴¹ (**e**; image adapted with permission from ref. ⁴⁷, American Physical Society). Here, the symbols display the experimental results and the full red line represents results of a purely geometric model. **f**, Recently, unidirectional sliding under superlubric conditions was observed for graphene nanoribbons dragged on gold surfaces¹⁴.

chemical contamination may hinder superlubricity by introducing new energy dissipation channels and mechanical wear. Therefore, further scaling-up efforts involve several challenges that are yet to be overcome before the practical application of superlubricity can be realized. To this end, one first has to gain comprehensive understanding of the tribological processes occurring at nanoscale contacts.

Nanoscale superlubricity

As mentioned above, the theoretical predictions of Peyrard and Aubry²⁸ and Shinjo and Hirano²⁷ (later revisited by Consoli et al.⁴⁶), and the pioneering experiments of Dienwiebel et al.⁴¹, set the foundations for our present perception of nanoscale structural superlubricity. Specifically, in ref. ⁴¹, the sliding friction of nanoscale graphite flakes dragged across a pristine graphitic surface was measured as a function of the flake/substrate relative angular orientation (see Fig. 2d, e). The measured friction was found to be vanishingly small (below experimental error) throughout the range of misfit angles studied except for narrow regions, of 60° periodicity, that exhibited stick-slip motion with substantially increased friction. This friction enhancement was associated with an aligned contact, in which the lattice vectors of the contacting hexagonal graphene layers forming the interface become parallel. Notably, the width of these friction peaks was found to reduce with increasing contact size, mainly due to better cancellation of the lateral forces that act on the flake atoms resulting in averaging-out of the net friction force^{47–49}.

The directional character of superlubricity in graphitic contacts was further demonstrated in free sliding experiments, in which nanoscale graphene flakes were pushed out of their optimal commensurate stacking configuration and allowed to slide above a graphene surface⁵⁰. In their superlubric state, the flakes were found to slide freely up to 100 nm at a temperature of 5 K, and were stopped only by their realignment with the underlying substrate. Owing to the enhancement of thermally induced reorientation processes, the free-sliding length decreased with increasing temperature.

Extending the scope of super-low friction studies beyond rigid layered material interfaces, the motion of graphene nanoribbons across Au(111) surfaces has been recently studied (see Fig. 2f)¹⁴. In this case, extremely small frictional forces, of the order of piconewtons, were

measured for graphene nanoribbons pulled along the $[-1, 0, 1]$ direction of the underlying gold surface. The forces were found to be nearly independent of the length of the nanoribbons, indicating superlubric motion of the incommensurate contact with residual friction forces originating from edge effects.

It was further demonstrated that superlubricity is not limited to fully crystalline contacts, but can also be realized in incommensurate junctions consisting of crystalline and disordered materials. In particular, the friction between a graphite surface and amorphous antimony or crystalline gold clusters was measured⁵¹. Within the context of the scale-up of superlubricity, the main focus of these studies was devoted to investigating the contact size dependence of the kinetic friction force. The traditional tribological view suggests a classical linear dependence, as often observed in the macroscopic world. This, indeed, is the case for commensurate nanoscale friction contacts²⁵. Remarkably, for clean contacts between the amorphous clusters and graphite the friction force was found to scale with the square root of the contact area. This scaling was rationalized by the fact that the lateral atomic forces in disordered surfaces average-out, following the central limit theorem⁵². In misaligned crystalline contacts between gold and graphite, where cancellation of the lateral forces results from the lattice incommensurability, even lower scaling exponents of the friction force with flake size were measured. It was further argued that in this case the exponent value is not unique but rather depends on the contact shape and scan-line direction^{51,53,54}. Interestingly, recent computational studies predicted that under certain conditions the frictional behaviour of incommensurate layered material contacts can become completely independent of the contact size^{53,55,56}. This finding suggests the intriguing possibility of scaling-up structural superlubricity towards the micro- and even macroscales.

Notably, similar effects of friction reduction due to inter-lattice commensuration effects have also been observed for interfaces of soft materials such as colloidal suspensions⁵⁷. The microscale dimensions of the individual colloids used in such set-ups allows for direct observation and investigation of the mechanisms underlying the transition from stick-slip to superlubric sliding. Furthermore, they facilitate explicit control over inter-particle and particle/substrate interaction parameters, thus enabling various frictional regimes to be explored. Apart

from their fundamental importance, such studies, combined with large-scale realistic simulations of appropriate model systems⁵⁸, shed light on important factors that govern friction and wear in crystalline contacts.

These striking examples of the strong interplay between experiment and theory are characteristic of the field of nanoscience in general and of nanotribology in particular. At such reduced length scales, theory and computation can provide highly reliable description of the physical processes underlying the measured phenomena. Hence, they can help in both the analysis and the rationalization of experimental observations and in the prediction of novel material behaviours. To this end, theory offers a spectrum of approaches, ranging from coarse-grained descriptions relying on geometric considerations to fully atomistic elaborate simulations. These allow deep understanding of nanoscale tribological effects to be gained and rational deductions about micro- and macroscale contacts to be made.

As discussed above, when considering wearless friction in rigid crystalline interfaces, the friction is found to be strongly related to the inter-lattice commensurability^{39–41,59}. In such cases, substantial insights can be gained from simple geometric descriptions of the contact that quantify the degree of lattice registry^{60,61}. As an example, in Fig. 2e we show that accounting for geometric considerations by using the registry index approach (red line) can capture the variation of the measured sliding friction with misfit angle (black points) in graphitic contacts⁶⁰. However, such descriptions neglect important effects such as in- and out-of-plane elasticity of the substrate and slider, dynamic reorientations, thermal fluctuations, energy dissipation processes, effects of chemical contact contamination, and possible wear. In principle, all these effects can be captured by *ab initio* molecular dynamics simulations, but these are prohibitively computationally expensive even for nanoscale contact models. Therefore, one often resorts to fully atomistic classical molecular mechanics simulations that rely on dedicated force fields, specifically parameterized to capture the corresponding intra- and inter-layer interactions^{62–68}. Even this approach is restricted by the simulated timescales limiting the calculation to interfacial shear velocities that are orders of magnitude higher than in typical friction force experiments. Here, semi-phenomenological approaches that focus on reduced dynamics of few important degrees of freedom may help bridge the gap between the timescales of experiments and those of simulations⁶⁹. In particular, such approaches have shown that friction often exhibits a logarithmic dependence on velocity, thus justifying the validity of fully atomistic simulations in the study of tribological processes in nanoscale junctions.

One of the most important contributions of computational simulations to the field of nanotribology is the ability to identify mechanisms that eliminate superlubricity and suggest ways to overcome them. In this respect, an important extrinsic factor that may suppress superlubricity was found to be the incorporation of contaminants, such as chemical adsorbates and various nanoparticles, within the frictional junction. These often lead to pinning of incommensurate surfaces, resulting in the appearance of static friction and enhancement of kinetic friction^{51,70–72}. Surface heating may lead to contaminant desorption, thus recovering the bare surfaces and reducing the adsorbate-related friction^{51,73–75}. Such heating would be most effective as a pre-treatment applied to the exposed surfaces before the formation of the junction. An alternative *in situ* approach was further suggested, in which mechanical oscillations of the frictional contact lead to substantial decontamination of the interface, thus restoring its super-low friction characteristics^{76,77}.

The normal load experienced by the frictional junction constitutes another extrinsic factor limiting superlubricity. Obviously, above a certain (system-dependent) normal load, any contact should exhibit increased friction leading to enhanced wear. It is therefore desirable to identify conditions under which superlubricity can be sustained with practically applied loads. Computational studies have revealed that for incommensurate contacts the edge atoms of a finite sliding flake are most prone to the effects of an increased normal load. This, in turn, may lead to enhanced friction via their pinning to the underlying surface⁷⁸. The importance of such effects, however, was shown to reduce with increasing contact size^{56,75}.

Furthermore, most experiments demonstrating superlubricity to date have been performed at relatively low driving velocities, of the order of micrometres per second, whereas practical applications, such as mechanical hard-disk drive read/write components (see below), often operate at considerably higher velocities of tens of metres per second. This, in turn, may enhance energy dissipation and wear effects, resulting in the elimination of superlubricity. Microscale understanding of these processes may be gained by molecular dynamics simulations that are most suitable for describing frictional behaviour at such high velocities.

Frictional junctions also possess intrinsic properties that may eliminate superlubricity. For example, considering the results of Dienwiebel *et al.*⁴¹, one may conclude that in order to achieve superlubricity in practical applications it is sufficient to bring two rigid crystalline surfaces into incommensurate contact. Nevertheless, both experiments and simulations have shown that dynamic reorientations of the sliding surfaces tend to lock the system into its commensurate high-friction configuration, thus limiting the realization of superlubricity to short timescales⁴⁹. Furthermore, as discussed above, intrinsic layer elasticity has been theoretically predicted to increase friction in incommensurate contacts. Here the typical out-of-plane stiffness of the individual layers is comparable to the effective stiffness of the inter-layer shear force hence providing accessible energy dissipation channels that may enhance friction already in nanoscale contacts^{71,72,79–82}. Furthermore, while the corresponding in-plane stiffness is considerably higher, its effects are expected to already be manifested in microscale contacts, where elastic deformations may become sufficiently large⁸³ that commensurate regions can develop, leading to an increase in friction^{35,36,84}.

To address these issues, nanoscale heterogeneous junctions formed between graphite and hexagonal boron nitride⁸⁵ have been suggested theoretically to provide an accessible route towards robust superlubricity (see Fig. 1)^{56,86,87}. It was shown that, above a certain contact size, the corrugation of their sliding potential energy surface is considerably reduced even for aligned contacts. This effect, associated with the appearance of moiré patterns resulting from the intrinsic inter-layer lattice constant mismatch in heterojunctions, suggests that superlubricity should be preserved even under interfacial reorientations^{86,87}. It was further shown that superlubricity in heterogeneous interfaces of graphene and hexagonal boron nitride can sustain considerably higher loads than in their homogeneous graphitic counterparts⁸⁶. This was attributed to the intrinsic incommensurability that reduces the effects of edge atom pinning under high normal loads. Recent experimental evidence showing that heterogeneous microscale graphite and hexagonal boron nitride contacts exhibit superlubricity that is nearly orientationally independent supports that prediction⁸⁸. Interestingly, a related experiment⁴⁵ demonstrated similar behaviour for both heterogeneous graphene/hexagonal boron nitride and homogeneous graphitic contacts. Here, the polycrystalline nature of one of the contacting surfaces further prohibited the formation of fully commensurate contacts⁸⁹. Additional support was also provided by experimental observations of self-orientation of microscale hexagonal boron nitride flakes on graphitic surfaces⁹⁰.

Micrometre- and centimetre-scale superlubricity

At the forefront of the field of superlubricity stands the effort to demonstrate the robustness of the effect against intrinsic elimination mechanisms and external perturbations at ever-growing contact dimensions. Recently, two decades after the first experimental demonstration of nanoscale structural superlubricity, evidence of frictionless sliding in micrometre- and centimetre-sized crystalline contacts has been reported^{42–45,88}.

The first experimental observation of microscale structural superlubricity was reported in 2012⁴² (see Fig. 1). Clean single-crystalline interfaces between two graphitic stacks were formed via shear-force-induced mechanical exfoliation of a multilayer graphitic mesa (see Fig. 3a)⁴². The shear force was applied by an external tip to the mesa top causing a stacking fault at the weakest interface that divides it into

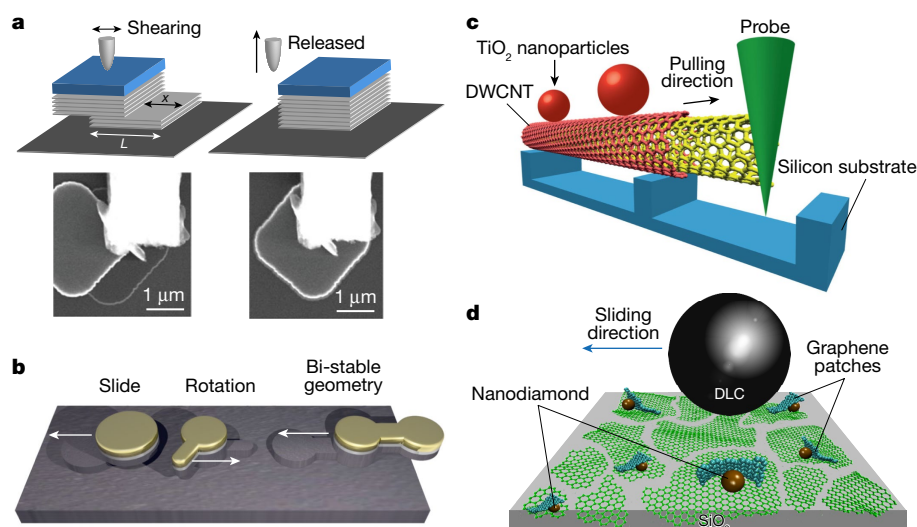


Fig. 3 | Microscale superlubricity. **a**, Self-retraction motion of a microscale square graphitic mesa. An initially sheared mesa (illustrated and imaged in the upper and lower left subpanels, respectively) self-retracts to its original position (illustrated and imaged in the upper and lower right subpanels, respectively) (image reprinted with permission from ref. ⁴², American Physical Society). **b**, Shear force measurement in mesoscopic graphite contacts (image reprinted with permission from

ref. ⁴⁴, American Association for the Advancement of Science). **c**, Inter-wall telescopic superlubric motion in centimetre-long double-walled carbon nanotubes (DWCNT; image reprinted with permission from ref. ¹⁰⁰, Springer Nature). **d**, Multi-contact superlubricity in microscopic interfaces between diamond-like carbon (DLC) and graphene scrolls wrapped around diamond nanoparticles (image adapted with permission from ref. ⁷, American Association for the Advancement of Science).

two weakly interacting graphite stacks. The upper stack could then be repeatedly sheared against the lower one in different directions and relative angular orientations. Notably, upon release, most sheared stacks returned to their original position with no external aid, exhibiting self-retraction due to an adhering restoring force that drives the system towards minimum interfacial energy⁹¹. For specific orientations, of six-fold rotational symmetry, a lock-in effect was demonstrated with no evident self-retraction. This clearly indicates that the microscale interface is constructed from single-crystalline graphene layers that exhibit superlubric self-retraction motion when placed at incommensurate configurations. Further support for this conclusion was recently provided⁷⁵, when quantitative measurements of the tribological properties of this system demonstrated dynamic friction coefficients well within the superlubric regime for the misaligned contact up to external normal loads of 1.67 MPa. Importantly, the superlubric behaviour was found to sustain not only vacuum conditions but also ambient conditions at various humidities⁴² and high sliding velocities up to 73 m s^{-1} . Additional experimental evidence of superlubricity in microscale graphitic contacts was recently provided when measurements of the dynamic friction force as a function of contact size yielded power-law scaling with a typical exponent of 0.35, lower than the value of 0.5 characteristic of amorphous contact, thus indicating the formation of an incommensurate crystalline contact (see Fig. 3b)⁴⁴. However, a major drawback of such microscale homogeneous graphitic junctions is that friction increases dramatically when the contacting surfaces are aligned. As mentioned above, recent theoretical predictions⁸⁶ and experiments⁸⁸ on graphene/hexagonal boron nitride contacts demonstrated that heterojunctions in layered materials may offer a remedy for this problem.

One of the most recent advances in the field of superlubricity involved the extension of the scope of superlubricity to the centimetre regime⁴³. This became possible by taking advantage of the intrinsic inter-wall incommensurability in pristine bichiral double-walled carbon nanotubes. An inner-shell pull-out experiment was used to measure the inter-wall friction in coaxial centimetre-long double-walled carbon nanotubes, yielding friction forces as low as 1 nN independent of the axial shift extension (see Fig. 3c). Such systems hold the technological potential to serve as low-energy dissipative gigahertz oscillators^{92,93}. We note that the actual contact area in this experiment was comparable to that of the self-retracting graphitic mesas discussed

above. Nevertheless, centimetre-scale single-crystalline graphene surfaces are already achievable today and hold great potential for large-scale tribological applications (see Fig. 4)⁹⁴. Furthermore, the fact that one of the system dimensions extends to the centimetre scale provides evidence that using judicious geometrical designs can help suppress intrinsic elimination mechanisms to allow superlubricity at large length scales.

Another promising route to obtain large-scale superlubricity involves multi-contact interfacial geometries (see Fig. 1). In this respect, it was recently demonstrated that ultralow friction coefficients can be achieved in mesoscale contacts formed between a diamond-like carbon sphere and a silicon dioxide surface covered by graphene patches and diamond nanoparticles (see Fig. 3d)⁷. Under shear, the graphene patches tend to scroll around the diamond nanoparticles forming an incommensurate contact with the surface of the diamond-like carbon sphere. As mentioned above, another realization of the multi-contact approach to microscale superlubricity was recently provided, in which the friction between a graphene coated microsphere and a silicon dioxide surface covered by graphene or hexagonal boron nitride was shown to be considerably lower than between the bare surfaces⁴⁵. The underlying mechanism relies on the roughness of the sphere surface that incorporates elevated asperities covered by randomly oriented graphene patches. These form multiple nanoscale contacts of different registry with the substrate, prohibiting the formation of a fully commensurate contact. Importantly, these examples demonstrated that the multi-contact approach, involving the cumulative effect of many nanoscale contacts that are randomly distributed on the surface, is less susceptible to effects of external loads and humidity.

These recent experimental demonstrations of superlubricity in microscale contacts constitute an important milestone on the way to achieving superlubricity in macroscale contacts, which is essential for practical applications (see Fig. 4).

Challenges and future directions

Superlubricity holds great technological promise for reducing energy loss and friction-induced wear in actual mechanical devices. However, fulfilling this potential requires the scaling-up of superlubricity to macroscopic contacts. To meet this challenge several issues have to be addressed, as follows.

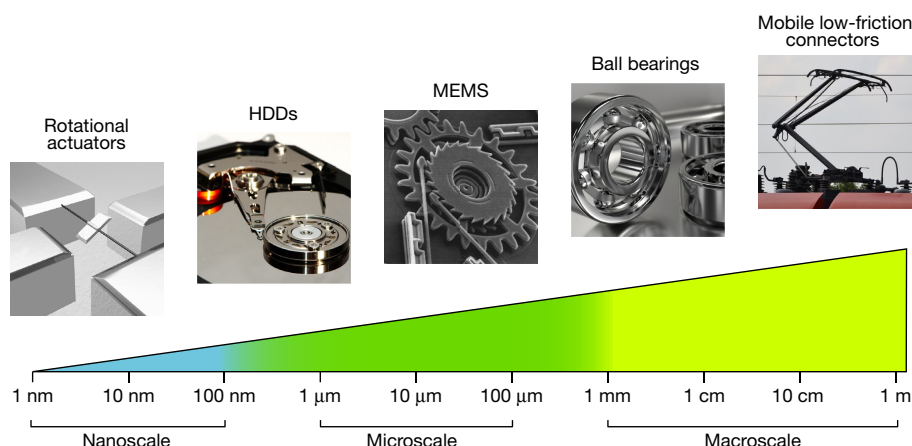


Fig. 4 | Demonstrative applications of structural superlubricity at different length scales. The bar is labelled underneath with the approximate positions of the length scales. Above the bar are images representing the applications, as follows (left to right): nanotube-based frictionless rotational actuators (image reproduced with permission from ref. ¹⁰¹, Springer Nature); wear-free nanoscale read/write contacts in hard-disk drives (HDDs; image reproduced with permission from Magnus

Hagdorn); durable micro-electro-mechanical systems (MEMS; image reproduced with permission from ref. ¹⁰², Annual Reviews); low-friction ball bearings; and efficient mobile low-friction connectors (image adapted from https://commons.wikimedia.org/wiki/File:Pantograph_of_a_DBAG_Class_423.JPG, published under a CC BY 3.0 DE licence (<https://creativecommons.org/licenses/by/3.0/de/deed.en>)).

Edge effects

Because the circumference-to-surface ratio of planar contacts decreases as the inverse contact dimension one could naively expect that edge effects would reduce with the interface size. However, considering the fact that the central contact area exhibits superlubric behaviour, edge pinning effects are often responsible for the residual friction and hence may become dominant with increasing contact size.

Surface roughness

In any practical application, layered materials will be deposited on substrates that will exhibit roughness at various length scales. When the surface corrugation becomes substantially larger than the typical length scales of the two-dimensional layered material crystal coat, friction may be dictated by the overall surface roughness rendering commensurability effects unimportant.

Surface defects and contaminants

Although pristine interfaces can be readily fabricated at the nanoscale, surface defects and contaminants are expected to appear with increasing contact size. Specifically, inter-layer covalent bonding and bulky molecular adsorbates that are present under ambient conditions may cause interfacial chemical pinning, which can considerably increase friction. Since the corresponding unbinding energy scales are typically very large even at low densities, such imperfections are expected to induce a considerable effect on the measured friction.

Elasticity

When incommensurate surfaces are put into contact, elasticity effects can support inter-lattice readjustment that leads to the formation of locally commensurate regions⁹⁵ characterized by strong lock-in forces. This effect grows with the contact size and hence is expected to enhance friction at macroscale junctions²⁵.

External conditions

Some of the less-explored factors in the field of superlubricity are the effects of external normal load, sliding velocity and distance, as well as temperature. Most experiments to date have considered normal loads and velocities that are substantially lower than those typically experienced in macroscale systems. It is expected that high loads and high sliding velocities will lead to enhanced wear of the two-dimensional material surface coating, which will reduce the interface durability⁹⁶. Furthermore, most measurements aim at understanding basic tribological phenomena and are thus performed at ambient temperature

and with very short sliding distances. In this respect, it should be noted that while under superlubric conditions the residual wear is expected to be low, it will not completely vanish as the friction coefficients remain finite. Hence, for practical applications experiments on superlubricity should consider large sliding distances as well as extreme temperatures and extreme normal loads.

On the basis of the substantial experimental and theoretical advances achieved over the past decade in the field of superlubricity, three promising routes can be suggested to overcome these challenges. First, the idea of obtaining robust structural superlubricity via the use of heterogeneous contacts of rigid layered materials can be extended to the macroscale. Here the intrinsic lattice misfit between the contacting surfaces leads to incommensurate positioning of the edge atoms that inhibits the formation of coherent edge pinning effects, thus considerably reducing the residual edge friction⁵⁶. In fact, structural superlubricity is not limited to the realm of layered materials and can be achieved under more general conditions. What is essential for the realization of the discussed mechanism is the formation of a contaminant-free interface between atomically smooth stiff surfaces that are out of registry. Moreover, the notion of disorder-induced incommensurability in heterojunctions formed between a crystal surface and an amorphous counterpart can also be extended to the macroscale. When the contacting surfaces are rigid—for example, diamond-like carbon—ultralow friction and wear can be achieved^{21,97}.

Second, multi-contact configurations can serve as a venue for resolving some of the above-mentioned problems, taking advantage of surface roughness and/or polycrystallinity to effectively transform macroscale junctions into a large collection of nanoscale contacts^{45,89}. Under appropriate conditions, where the various contacts are coated by randomly oriented patches of two-dimensional material, robust superlubricity is expected to prevail even under high external loads and high sliding velocities. Furthermore, multi-asperity configurations effectively decouple the individual nanoscale contacts, thus diminishing undesirable elasticity effects. Nevertheless, within this approach the effective contact edge-to-surface ratio considerably increases and that, in turn, may result in undesirable edge pinning effects and additional friction.

Third, a way to reduce in-plane elasticity effects would involve the deposition of two-dimensional material coatings on rigid surfaces and/or the usage of multi-layer stacks. The supporting bulk reduces the tendency of the contacting layers at the frictional interface to adjust their lattices, therefore diminishing the formation of locally commensurate regions and supporting superlubricity. Further reduction of inter-lattice adjustment effects may be achieved via extension or compression

stresses applied to the interfacing layers. Notably, in contrast to such in-plane elasticity effects, interfacial out-of-plane deformations may promote the occurrence of superlubricity via soliton-like smooth sliding of elevated moiré ridges that can extend up to the macroscale⁵⁶. Multi-layer systems may also increase interface rigidity and reduce roughness effects to some extent by eliminating surface rippling and decoupling the frictional interface from the underlying corrugated surface⁷¹. We note that the external normal loads that are often considered to enhance friction due to increased steric repulsions may also suppress surface rippling if applied uniformly to extended contacts. Therefore, one may expect to find regimes of negative friction coefficients, where the friction forces reduce with the external load due to flattening effects⁹⁸. Furthermore, similarly to the microscopic contact case^{76,77}, the use of external loads to induce lateral mechanical oscillations can serve as an interface pre-treatment procedure to dynamically eliminate surface contaminants.

In view of all of the above, we believe that to meet the challenge of achieving superlubricity in macroscopic contacts under realistic operation conditions one should adopt a synergistic strategy combining the advantages of the various approaches discussed herein. Of particular potential would be the use of multi-contact polycrystalline heterojunctions of clean multi-layer stacks. The combination of polycrystallinity and heterogeneous coatings is required to prohibit the formation of commensurate regions and reduce edge pinning effects over the entire interface. Multi-asperity configurations can harness the advantages of nanoscale contacts even at the macroscale and when underlying multi-layer stacks they can further eliminate undesirable elasticity effects. Finally, keeping such interfaces clean to avoid pinning effects should lead to robust macroscale superlubricity. Identifying material junctions that can satisfy these conditions will be a technological breakthrough that will revolutionize many engineering and industrial concepts and shape new paradigms in friction-induced energy loss and wear. The space, automotive and electronics industries, as well as medical manufacturers and information storage centres, among many others, may all benefit greatly from this forthcoming technology (see Fig. 4). In an age when natural resources are becoming limited and the environmental impact of their usage is affecting Earth's atmosphere and ecosystem, macroscale superlubricity may contribute to the reduction of both global energy consumption and pollutant emission.

Received: 6 February 2017; Accepted: 21 September 2018;
Published online 21 November 2018.

- Dowson, D. *History of Tribology* 2nd edn (Professional Engineering Publishing, London, 1998).
- Urbakh, M., Klafter, J., Gourdon, D. & Israelachvili, J. The nonlinear nature of friction. *Nature* **430**, 525–528 (2004).
- Bormuth, V., Varga, V., Howard, J. & Schäffer, E. Protein friction limits diffusive and directed movements of kinesin motors on microtubules. *Science* **325**, 870–873 (2009).
- Klein, J. Repair or replacement — a joint perspective. *Science* **323**, 47–48 (2009).
- Bhushan, B. *Principles and Applications of Tribology* 2nd edn (Wiley & Sons, New York, 2013).
- Holmberg, K., Andersson, P. & Erdemir, A. Global energy consumption due to friction in passenger cars. *Tribol. Int.* **47**, 221–234 (2012).
- Berman, D., Deshmukh, S. A., Sankaranarayanan, S. K. R. S., Erdemir, A. & Sumant, A. V. Macroscale superlubricity enabled by graphene nanoscroll formation. *Science* **348**, 1118–1122 (2015).
- Experimental demonstration of large-scale multi-contact structural superlubricity.**
- Granick, S. Motions and relaxations of confined liquids. *Science* **253**, 1374–1379 (1991).
- Granick, S., Zhu, Y. & Lee, H. Slippery questions about complex fluids flowing past solids. *Nat. Mater.* **2**, 221–227 (2003).
- Raviv, U. & Klein, J. Fluidity of bound hydration layers. *Science* **297**, 1540–1543 (2002).
- Briscoe, W. H. et al. Boundary lubrication under water. *Nature* **444**, 191–194 (2006).
- Lee, S. & Spencer, N. D. Sweet, hairy, soft, and slippery. *Science* **319**, 575–576 (2008).
- Pawlak, R. et al. Single-molecule tribology: force microscopy manipulation of a porphyrin derivative on a copper surface. *ACS Nano* **10**, 713–722 (2016).
- Kawai, S. et al. Superlubricity of graphene nanoribbons on gold surfaces. *Science* **351**, 957–961 (2016).
- Experimental demonstration of nanoscale structural superlubricity in graphene nanoribbons on gold surfaces.**
- Sweeney, J. et al. Control of nanoscale friction on gold in an ionic liquid by a potential-dependent ionic lubricant layer. *Phys. Rev. Lett.* **109**, 155502 (2012).
- Fajardo, O. Y., Bresme, F., Kornyshev, A. A. & Urbakh, M. Electro-tunable lubricity with ionic liquid nanoscale films. *Sci. Rep.* **5**, 7698 (2015).
- Rapoport, L. et al. Hollow nanoparticles of WS₂ as potential solid-state lubricants. *Nature* **387**, 791–793 (1997).
- Skinner, J., Gane, N. & Tabor, D. Micro-friction of graphite. *Nature* **232**, 195–196 (1971).
- Mate, C. M., McClelland, G. M., Erlandsson, R. & Chiang, S. Atomic-scale friction of a tungsten tip on a graphite surface. *Phys. Rev. Lett.* **59**, 1942–1945 (1987).
- Erdemir, A., Eryilmaz, O. L. & Fenske, G. Synthesis of diamondlike carbon films with superlow friction and wear properties. *J. Vac. Sci. Technol. A* **18**, 1987–1992 (2000).
- Erdemir, A. & Donner, C. Tribology of diamond-like carbon films: recent progress and future prospects. *J. Phys. D* **39**, R311–R327 (2006).
- Rozman, M. G., Urbakh, M. & Klafter, J. Controlling chaotic frictional forces. *Phys. Rev. E* **57**, 7340–7343 (1998).
- Socoliuc, A. et al. Atomic-scale control of friction by actuation of nanometer-sized contacts. *Science* **313**, 207–210 (2006).
- Lantz, M. A., Wiesmann, D. & Gotsmann, B. Dynamic superlubricity and the elimination of wear on the nanoscale. *Nat. Nanotechnol.* **4**, 586–591 (2009).
- Müser, M. H. Structural lubricity: role of dimension and symmetry. *Europhys. Lett.* **66**, 97–103 (2004).
- Considerations of elasticity and contact dimension for structural superlubricity.**
- Erdemir, A. & Martin, J.-M. (eds) *Superlubricity* (Elsevier, Amsterdam, 2007).
- Shinjo, K. & Hirano, M. Dynamics of friction — superlubric state. *Surf. Sci.* **283**, 473–478 (1993).
- First computational study of crystalline commensuration effects on kinetic friction reduction and coining of the term 'superlubricity'.**
- Peyrard, M. & Aubry, S. Critical behaviour at the transition by breaking of analyticity in the discrete Frenkel-Kontorova model. *J. Phys. C* **16**, 1593–1608 (1983).
- First theoretical prediction of elimination of static friction in incommensurate contacts.**
- Hirano, M., Shinjo, K., Kaneko, R. & Murata, Y. Anisotropy of frictional forces in muscovite mica. *Phys. Rev. Lett.* **67**, 2642–2645 (1991).
- Prandtl, L. Ein Gedankenmodell zur kinetischen Theorie der festen Körper. *Z. Angew. Math. Mech.* **8**, 85–106 (1928).
- Tomlinson, G. A. CVI. A molecular theory of friction. *Lond. Edinb. Dublin Phil. Mag. J. Sci.* **7**, 905–939 (1929).
- Frenkel, Y. & Kontorova, T. On the theory of plastic deformation and twinning. *Phys. Z. Sowjetunion* **13**, 1–10 (1938).
- Bylinskii, A., Gangloff, D., Counts, I. & Vuletic, V. Observation of Aubry-type transition in finite atom chains via friction. *Nat. Mater.* **15**, 717–721 (2016).
- Sørensen, M. R., Jacobsen, K. W. & Stoltze, P. Simulations of atomic-scale sliding friction. *Phys. Rev. B* **53**, 2101–2113 (1996).
- Ma, M., Benassi, A., Vanossi, A. & Urbakh, M. Critical length limiting superlow friction. *Phys. Rev. Lett.* **114**, 055501 (2015).
- Sharp, T. A., Pastewka, L. & Robbins, M. O. Elasticity limits structural superlubricity in large contacts. *Phys. Rev. B* **93**, 121402 (2016).
- Dietzel, D., Brndiar, J., Štich, I. & Schirmeisen, A. Limitations of structural superlubricity: chemical bonds versus contact size. *ACS Nano* **11**, 7642–7647 (2017).
- Blatter, G., Feigel'man, M. V., Geshkenbein, V. B., Larkin, A. I. & Vinokur, V. M. Vortices in high-temperature superconductors. *Rev. Mod. Phys.* **66**, 1125–1388 (1994).
- Martin, J. M., Donnet, C., Le Mogne, T. & Epicer, T. Superlubricity of molybdenum disulphide. *Phys. Rev. B* **48**, 10583–10586 (1993).
- Sheehan, P. E. & Lieber, C. M. Nanotribology and nanofabrication of MoO₃ structures by atomic force microscopy. *Science* **272**, 1158–1161 (1996).
- Dienwiebel, M. et al. Superlubricity of graphite. *Phys. Rev. Lett.* **92**, 126101 (2004).
- First experimental demonstration of structural superlubricity at nanoscale layered material contacts.**
- Liu, Z. et al. Observation of microscale superlubricity in graphite. *Phys. Rev. Lett.* **108**, 205503 (2012).
- First experimental demonstration of structural superlubricity in microscale graphitic contacts.**
- Zhang, R. et al. Superlubricity in centimetres-long double-walled carbon nanotubes under ambient conditions. *Nat. Nanotechnol.* **8**, 912–916 (2013).
- Koren, E., Lörtscher, E., Rawlings, C., Knoll, A. W. & Duerig, U. Adhesion and friction in mesoscopic graphite contacts. *Science* **348**, 679–683 (2015).
- Experimental demonstration of structural superlubricity in mesoscale graphitic contacts.**
- Liu, S.-W. et al. Robust microscale superlubricity under high contact pressure enabled by graphene-coated microsphere. *Nat. Commun.* **8**, 14029 (2017).
- Consoli, L., Knops, H. J. F. & Fasolino, A. Onset of sliding friction in incommensurate systems. *Phys. Rev. Lett.* **85**, 302–305 (2000).
- Hod, O. Interlayer commensurability and superlubricity in rigid layered materials. *Phys. Rev. B* **86**, 075444 (2012).
- Verhoeven, G. S., Dienwiebel, M. & Frenken, J. W. M. Model calculations of superlubricity of graphite. *Phys. Rev. B* **70**, 165418 (2004).

49. Filippov, A. E., Dienwiebel, M., Frenken, J. W. M., Klafter, J. & Urbakh, M. Torque and twist against superlubricity. *Phys. Rev. Lett.* **100**, 046102 (2008).
Theoretical and experimental investigation of elimination of structural superlubricity due to contact reorientations.
50. Feng, X., Kwon, S., Park, J. Y. & Salmeron, M. Superlubric sliding of graphene nanoflakes on graphene. *ACS Nano* **7**, 1718–1724 (2013).
51. Dietzel, D., Feldmann, M., Schwarz, U. D., Fuchs, H. & Schirmeisen, A. Scaling laws of structural lubricity. *Phys. Rev. Lett.* **111**, 235502 (2013).
52. Müser, M. H. in *Fundamentals of Friction and Wear on the Nanoscale* (eds Gnecco, E. & Meyer, E.) 177–199 (Springer, Switzerland, 2007).
53. de Wijn, A. S. (In)commensurability, scaling, and multiplicity of friction in nanocrystals and application to gold nanocrystals on graphite. *Phys. Rev. B* **86**, 085429 (2012).
54. Koren, E. & Duerig, U. Moiré scaling of the sliding force in twisted bilayer graphene. *Phys. Rev. B* **94**, 045401 (2016).
55. Koren, E. & Duerig, U. Superlubricity in quasicrystalline twisted bilayer graphene. *Phys. Rev. B* **93**, 201404 (2016).
56. Mandelli, D., Leven, I., Hod, O. & Urbakh, M. Sliding friction of graphene/hexagonal-boron nitride heterojunctions: a route to robust superlubricity. *Sci. Rep.* **7**, 10851 (2017).
57. Bohlein, T., Mikhael, J. & Bechinger, C. Observation of kinks and antikinks in colloidal monolayers driven across ordered surfaces. *Nat. Mater.* **11**, 126–130 (2012).
58. Vanossi, A., Manini, N. & Tosatti, E. Static and dynamic friction in sliding colloidal monolayers. *Proc. Natl Acad. Sci. USA* **109**, 16429–16433 (2012); correction **109**, 20774 (2012).
59. Li, H. et al. Superlubricity between MoS₂ monolayers. *Adv. Mater.* **29**, 1701474 (2017).
60. Hod, O. The registry index: a quantitative measure of materials' interfacial commensurability. *ChemPhysChem* **14**, 2376–2391 (2013).
61. Ward, M. D. Soft crystals in flatland: unraveling epitaxial growth. *ACS Nano* **10**, 6424–6428 (2016).
62. Tersoff, J. Empirical interatomic potential for carbon, with applications to amorphous carbon. *Phys. Rev. Lett.* **61**, 2879–2882 (1988).
63. Brenner, D. W. Tersoff-type potentials for carbon, hydrogen and oxygen. *Mater. Res. Soc. Symp. Proc.* **141**, 59–64 (1988).
64. Kolmogorov, A. N. & Crespi, V. H. Smoothest bearings: interlayer sliding in multiwalled carbon nanotubes. *Phys. Rev. Lett.* **85**, 4727–4730 (2000).
65. Kolmogorov, A. N. & Crespi, V. H. Registry-dependent interlayer potential for graphitic systems. *Phys. Rev. B* **71**, 235415 (2005).
66. Leven, I., Azuri, I., Kronik, L. & Hod, O. Inter-layer potential for hexagonal boron nitride. *J. Chem. Phys.* **140**, 104106 (2014).
67. Leven, I., Guerra, R., Vanossi, A., Tosatti, E. & Hod, O. Multiwalled nanotube faceting unravelled. *Nat. Nanotechnol.* **11**, 1082–1086 (2016).
68. Leven, I., Maaravi, T., Azuri, I., Kronik, L. & Hod, O. Interlayer potential for graphene/h-BN heterostructures. *J. Chem. Theory Comput.* **12**, 2896–2905 (2016).
69. Vanossi, A., Manini, N., Urbakh, M., Zapperi, S. & Tosatti, E. Colloquium: Modeling friction: from nanoscale to mesoscale. *Rev. Mod. Phys.* **85**, 529–552 (2013).
70. Müser, M. H., Wenning, L. & Robbins, M. O. Simple microscopic theory of Amontons's laws for static friction. *Phys. Rev. Lett.* **86**, 1295–1298 (2001).
71. Lee, C. et al. Frictional characteristics of atomically thin sheets. *Science* **328**, 76–80 (2010).
72. Zheng, X. et al. Robust ultra-low-friction state of graphene via moiré superlattice confinement. *Nat. Commun.* **7**, 13204 (2016).
73. Yang, J. et al. Observation of high-speed microscale superlubricity in graphite. *Phys. Rev. Lett.* **110**, 255504 (2013).
74. Wang, W. et al. Measurement of the cleavage energy of graphite. *Nat. Commun.* **6**, 7853 (2015).
75. Vu, C. C. et al. Observation of normal-force-independent superlubricity in mesoscopic graphite contacts. *Phys. Rev. B* **94**, 081405 (2016).
76. Liu, Z. et al. A graphite nanoeraser. *Nanotechnology* **22**, 265706 (2011).
77. Ma, M. et al. Diffusion through bifurcations in oscillating nano- and microscale contacts: fundamentals and applications. *Phys. Rev. X* **5**, 031020 (2015).
78. van Wijk, M. M., Dienwiebel, M., Frenken, J. W. M. & Fasolino, A. Superlubric to stick-slip sliding of incommensurate graphene flakes on graphite. *Phys. Rev. B* **88**, 235423 (2013).
79. Bonelli, F., Manini, N., Cadelano, E. & Colombo, L. Atomistic simulations of the sliding friction of graphene flakes. *Eur. Phys. J. B* **70**, 449–459 (2009).
80. Reguzzoni, M., Fasolino, A., Molinari, E. & Righi, M. C. Friction by shear deformations in multilayer graphene. *J. Phys. Chem. C* **116**, 21104–21108 (2012).
81. Gao, W. & Tkatchenko, A. Sliding mechanisms in multilayered hexagonal boron nitride and graphene: the effects of directionality, thickness, and sliding constraints. *Phys. Rev. Lett.* **114**, 096101 (2015).
82. Ouyang, W., Ma, M., Zheng, Q. & Urbakh, M. Frictional properties of nanojunctions including atomically thin sheets. *Nano Lett.* **16**, 1878–1883 (2016).
83. Annett, J. & Cross, G. L. W. Self-assembly of graphene ribbons by spontaneous self-tearing and peeling from a substrate. *Nature* **535**, 271–275 (2016).
84. Kim, W. K. & Falk, M. L. Atomic-scale simulations on the sliding of incommensurate surfaces: the breakdown of superlubricity. *Phys. Rev. B* **80**, 235428 (2009).
85. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499**, 419–425 (2013).
86. Leven, I., Krepel, D., Shemesh, O. & Hod, O. Robust superlubricity in graphene/h-BN heterojunctions. *J. Phys. Chem. Lett.* **4**, 115–120 (2013).
Theoretical prediction of robust structural superlubricity in layered material heterojunctions.
87. Ansari, N., Nazari, F. & Illas, F. Role of structural symmetry breaking in the structurally induced robust superlubricity of graphene and h-BN homo- and hetero-junctions. *Carbon* **96**, 911–918 (2016).
88. Song, Y. et al. Robust microscale superlubricity in graphite/hexagonal boron nitride layered heterojunctions. *Nat. Mater.* **17**, 894–899 (2018).
89. de Wijn, A. S., Fasolino, A., Filippov, A. E. & Urbakh, M. Low friction and rotational dynamics of crystalline flakes in solid lubrication. *Eur. Phys. Lett.* **95**, 66002 (2011).
90. Woods, C. R. et al. Macroscopic self-reorientation of interacting two-dimensional crystals. *Nat. Commun.* **7**, 10800 (2016).
91. Zheng, Q. et al. Self-retracting motion of graphite microflakes. *Phys. Rev. Lett.* **100**, 067205 (2008).
92. Cumings, J. & Zettl, A. Low-friction nanoscale linear bearing realized from multiwall carbon nanotubes. *Science* **289**, 602–604 (2000).
93. Zheng, Q. & Jiang, Q. Multiwalled carbon nanotubes as gigahertz oscillators. *Phys. Rev. Lett.* **88**, 045503 (2002).
94. Lee, J.-H. et al. Wafer-scale growth of single-crystal monolayer graphene on reusable hydrogen-terminated germanium. *Science* **344**, 286–289 (2014).
95. Wijk, M. M. v., Schuring, A., Katsnelson, M. I. & Fasolino, A. Relaxation of moiré patterns for slightly misaligned identical lattices: graphene on graphite. *2D Mater.* **2**, 034010 (2015).
96. Klemen, A. et al. Atomic scale mechanisms of friction reduction and wear protection by graphene. *Nano Lett.* **14**, 7145–7152 (2014).
97. Wang, Y., Guo, J., Zhang, J. & Qin, Y. Ultralow friction regime from the in situ production of a richer fullerene-like nanostructured carbon in sliding contact. *RSC Advances* **5**, 106476–106484 (2015).
98. Deng, Z., Smolyanitsky, A., Li, Q., Feng, X.-Q. & Cannara, R. J. Adhesion-dependent negative friction coefficient on chemically modified graphite at the nanoscale. *Nat. Mater.* **11**, 1032–1037 (2012).
99. Maaravi, T., Leven, I., Azuri, I., Kronik, L. & Hod, O. Interlayer potential for homogeneous graphene and hexagonal boron nitride systems: reparametrization for many-body dispersion effects. *J. Phys. Chem. C* **121**, 22826–22835 (2017).
100. Urbakh, M. Friction: towards macroscale superlubricity. *Nat. Nanotechnol.* **8**, 893–894 (2013).
101. Fennimore, A. M. et al. Rotational actuators based on carbon nanotubes. *Nature* **424**, 408–410 (2003).
102. Sniegowski, J. J. & de Boer, M. P. IC-compatible polysilicon surface micromachining. *Annu. Rev. Mater. Sci.* **30**, 299–333 (2000).

Acknowledgements O.H. is grateful to the Israel Science Foundation (grant no. 1586/17), the Lise-Meitner Minerva Center for Computational Quantum Chemistry, the Center for Nanoscience and Nanotechnology at Tel-Aviv University, and The Naomi Foundation for their financial support. E.M. acknowledges support from the Swiss National Science Foundation, the Swiss Nanoscience Institute and COST Action MP1303. Q.Z. acknowledges financial support from NSFC (grant no. 1127202, 1147215), the National Basic Research Program of China (grant nos 2013CB934200 and 2010CB631005), SRFDP (grant no. 20130002110043), and the Cyrus Tang Foundation. M.U. acknowledges financial support from the Israel Science Foundation (grant no. 1141/18), COST Action MP1303, and the Center for Nanoscience and Nanotechnology at Tel-Aviv University. We thank E. Koren and A. Erdemir for sharing high-resolution versions of Fig. 3b and d, respectively.

Reviewer information *Nature* thanks A. Erdemir, E. Riedo, A. Schirmeisen, S. Zapperi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions O.H., E.M., Q.Z. and M.U. conceived the idea of writing this Perspective, devised its general structure, designed the figures, and contributed to the writing.

Competing interests The authors declare no competing interests.

Additional information

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to O.H. or Q.Z.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

LEDs for photons, physiology and food

P. M. Pattison^{1*}, J. Y. Tsao², G. C. Brainard³ & B. Bugbee⁴

Lighting based on light-emitting diodes (LEDs) not only is more energy efficient than traditional lighting, but also enables improved performance and control. The colour, intensity and distribution of light can now be controlled with unprecedented precision, enabling light to be used both as a signal for specific physiological responses in humans and plants, and as an efficient fuel for fresh food production. Here we show how a broad and improved understanding of the physiological responses to light will facilitate greater energy savings and provide health and productivity benefits that have not previously been associated with lighting.

Light is central to the biological history of our planet, both as a fuel for photosynthesis and as an environmental signal. As a fuel for photosynthesis light produces adenosine triphosphate, the universal energy currency of plants and animals. Sunlight powered the life whose fossilized remains have been human energy currency for the past two centuries, and it powers the photovoltaic- and wind-generated electricity that will become the energy sources of the future. As a signaller, light carries much of the information that enables life to adapt to its environment, and improved ability to receive that information is responsible for numerous evolutionary adaptations^{1,2}. The importance of visible-light signalling for humans is demonstrated by three observations: the exquisiteness of the eye as an optical instrument; the large fraction (half) of the human brain devoted to visual signal processing; and our extreme dependence on vision technologies, such as eyeglasses³. Light also regulates human circadian, neuroendocrine and neurobehavioural physiology^{4,5}, and has an even greater effect on plants, which have more photoreceptors to process light than do humans. The importance of light as fuel and as a signaller rendered traditional lighting essential to human civilization for basic illumination; LED lighting, with its greater level of engineering control, might trigger a new world of applications.

Light for basic illumination

Lighting was among the earliest of human technologies. It expands the productive day into non-sunlit hours⁶, and during the day it expands the productive space into the non-sunlit areas of enclosed spaces⁷. As illustrated in Fig. 1a, lighting technology has undergone successive fundamental improvements over the centuries, from chemical-fuel-based to vacuum-based electric lighting, culminating now in LED-based solid-state lighting^{8,9}.

As short an historical time as electric lighting has been available, on the horizon now are the outlines of a superior technology for basic illumination: LED-based solid-state lighting. The phosphor-converted light-emitting diode (PC-LED) approach is currently the most prevalent. A highly efficient blue LED is combined with optical down-converters, typically phosphors, which absorb a portion of the blue light and emit longer wavelengths to produce white. PC-LED white lighting has made such great progress in terms of lighting performance, efficiency and cost that there is little doubt that it will soon become the source of almost all electric lighting. This progress, which was triggered by foundational breakthroughs in the synthesis of AlInGaN semiconductors (for which the Nobel Prize in Physics was awarded¹⁰ in 2014), is illustrated¹¹ in Fig. 1b. Among the key advances were the uniform and controlled epitaxial growth of InGaN LEDs by metal-organic chemical vapour deposition;

thin-film flip-chip¹² and other device designs with high electrical and optical efficiencies; and robust high-quantum-yield phosphors emitting in the green-yellow and red regions of the spectrum.

LED lighting products not only use less energy and have a lower cost of ownership than other lighting technologies, but they can also possess other features that are of importance to human use of basic white-light illumination: desirable colour qualities (for example, colour rendering index and correlated colour temperature), minimal or no flicker, long life, and negligible environmental and human toxicity. The three main benefits of LED lighting are therefore a decrease in electricity consumption, a reduced cost of ownership, and improvements in lighting quality. The decrease in electricity consumption is huge because human society consumes so much electric light: around 6.5% of total global primary energy¹³ was used for lighting in 2005. It is forecast¹⁴ that in the United States alone LEDs will penetrate around 86% of electrical lighting installations by 2035, decrease electricity consumption for lighting by around 75% and save approximately 5.1 quadrillion British thermal units (5.1 quads) per year (around US\$52 billion per year) in direct energy cost.

Engineered light

Even with this remarkable progress, LED lighting is only in its infancy. The historic purpose of all lighting technologies has been to provide basic illumination for visibility and vision. We now stand at the threshold of what might be called 'engineered light', in which building blocks of solid-state lighting are combined for integrated functionality beyond basic human illumination^{15,16}. Four features will be especially important.

The first is spectral control. LED lighting originates in efficient narrowband blue-light emission from a semiconductor LED. This blue light can be combined with green-yellow and red optical down-converters to create various hues of white light, or with other direct-emitting LEDs (Fig. 2a, c). For the first time in history, spectral content can be customized for a wide range of applications, each with its own action spectrum. For basic illumination, the spectra can be engineered to match the human photopic visual response (the solid green curve in Fig. 2b). It can also be engineered for qualities including colour rendition, colour gamut and correlated colour temperature, and thus for controlled rendering of colours and various human visual preferences. For human health and productivity, as discussed later in this Perspective, the spectra can be engineered—as illustrated by the intrinsically photosensitive retinal ganglion cell bodies (ipRGCs) curve in Fig. 2b—for potent stimulation of ocular photoreceptors that regulate human circadian, neuroendocrine and neurobehavioural responses.

¹SSLs, Inc., Johnson City, TN, USA. ²Sandia National Laboratories, Albuquerque, NM, USA. ³Thomas Jefferson University, Philadelphia, PA, USA. ⁴Utah State University, Logan, UT, USA.
*e-mail: morgan@sslsinc.com

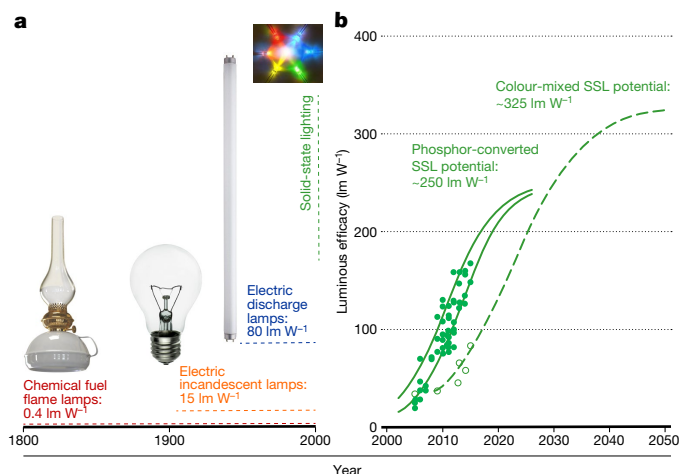


Fig. 1 | The history of lighting technology. **a**, The history of lighting. Chemical-fuel-based lighting was the earliest lighting technology. This was followed by electric lighting, which ushered in the modern era of electricity. Incandescent lamps, so widespread by the early twentieth century that they have now been taken for granted for generations, are an order of magnitude brighter and more efficient than gas lights. The incandescent lamp greatly reduced the volume of material that heats and emits light, thus enabling extremely high temperatures, shifting the spectrum of the black-body radiation from the infrared towards the visible. Electric discharge lamps, based on light emission from excited electronic states in controlled gas plasmas, are even more efficient—five times more so than incandescent lamps. Solid-state lighting reaches higher efficiencies still, with substantial increases expected in the future. Photo of six-colour LEDs courtesy of E. F. Schubert. **b**, Historical and projected luminous efficacies of solid-state lighting. In the past decade, luminous efficacy has tripled, from 40–60 lm W^{-1} to 140–160 lm W^{-1} , with even larger decreases in cost; in the coming decade, 325 lm W^{-1} is potentially achievable. Filled circles and solid lines, phosphor-converted LEDs; open circles and dashed line, colour-mixed LEDs; data from ref. ¹⁹.

For plants, also discussed later, the spectra can be engineered to stimulate photobiological responses that alter plant shape, increase photosynthesis and enhance nutritional value (Fig. 2d).

The second important feature of LED lighting is the precise control of the intensity of the light. Semiconductor LEDs are current-driven devices, the intensity of which can be precisely controlled across their operating range and modulated over a large range of frequencies. At the high (GHz) end of these rates, the modulation can be used for free-space visible-light communication (Li-Fi), possibly to alleviate the congestion and bandwidth limitations of Wi-Fi. In the medium range (kHz), pulse-width-modulation-controlled flicker-free dimming could help to match in real time the brightness of an illuminated scene with human visual preference. At the low (seconds–hours–days) end of these rates, modulation can be used to ensure that photons are emitted only when human eyes are available to perceive them, or to match natural, diurnal or seasonal lighting conditions.

The third feature is the control of distribution in space. Semiconductor LEDs emit light from extremely small areas and have low étendue. As such, they can be optically imaged in space with great precision or coupled efficiently into transparent waveguides with complex light-scattering surfaces. Semiconductor LEDs can also be easily arrayed and, through individual addressing and control, create pixelated ‘super beams’, the shapes of which are digitally controlled like a projection display.

The fourth important feature of LED lighting is its ready integration with other technologies. At the chip-and-package level, examples could include semiconductor technologies such as drivers; wireless and wired communication chips; photo, image, chemical, temperature or humidity sensors; and microprocessor and memory chips for local intelligence. At the luminaire level, other technologies could include acoustic transducers (microphones and speakers), radar- and lidar-based

three-dimensional scene mappers, and occupancy sensors to enhance the functionality of the lighting products. Perhaps most importantly, such integration enables lighting that is ‘connected’ to the Internet of Things. Connected lighting makes use of luminaires and light fixtures as the most ubiquitous of all grid-connected appliances. Connectivity enhances the fundamental benefits of LED lighting by enabling lights to sense and respond to their environment and communicate the conditions or their status. Having sensors available everywhere a light fixture is available could allow for an unprecedented degree of spatially and temporally resolved information communicated to the Internet of Things¹⁷, enabling new applications and advancing human productivity and safety. Even within a single building, connected controls can provide information on room utilization, improving our ability to load-schedule heating, cooling, lighting and other appliances in buildings. Such load scheduling will only grow in importance as the world’s energy economy continues to electrify via renewable but intermittent sources (for example, solar and wind)¹⁸.

These four features are catalysing a new world of engineered lighting that goes well beyond basic illumination. Although not without challenges¹⁹, research in laboratories worldwide will continue to make these features more powerful and more affordable. They can then be combined in new ways to add value to society that is even greater than just the energy savings. We highlight two of the most important ways below.

Lighting for human health and productivity

Basic illumination enables humans to see via their primary optical tract, and hence permits productivity indoors and/or at night when sunlight is unavailable. It is now known that the primary optical tract is only one of two photoreceptor pathways between the eye and the brain. The second, illustrated in Fig. 3, is the retinohypothalamic tract, which has a primary role in supporting the light regulation of human circadian, neurobehavioural and neuroendocrine responses^{4,5,20}, and ultimately impacts human health and productivity.

Scientists have only recently been able to delineate the photoreceptive input to the circadian and neuroendocrine systems. In 2001, two analytical action spectra identified 446–477 nm as the most potent region for acute melatonin suppression in healthy human subjects^{21,22}. Other complete analytical action spectra and studies that used selected-wavelength comparisons further indicated that circadian-phase shifting, autonomic stimulation, and the acute effects of light on alertness and performance are shifted towards the shorter wavelength—or blue-appearing—part of the visible spectrum^{4,5,23}. Taken together, these results indicated that a novel ocular photosensory system, distinct from the canonical rods and cones of the visual system, is primarily responsible for regulating physiology and behaviour in humans.

At the front end of the retinohypothalamic tract is a small population of widely dispersed ipRGCs that are directly responsive to light via a vitamin A photopigment named melanopsin^{24–28}. These then project through the retinohypothalamic tract to the paired hypothalamic suprachiasmatic nuclei, as well as to a number of other nuclei involved in regulating physiology and behaviour^{25,27,29}. The suprachiasmatic nuclei are master oscillators in the circadian system that transmit information about lighting and circadian time to diverse loci in the nervous system, including to the pineal gland where the hormone melatonin is synthesized^{4,5}. This circadian pacemaker thus synchronizes the systems of sleep- and arousal-promoting neurons in the central nervous system, and in turn the daily rhythms of sleep and wakefulness, body temperature, alertness, psychomotor performance, neurocognitive responses, and the secretion of hormones such as melatonin and cortisol^{4,5,20,30}.

In humans, the alerting effects of light have been assessed by subjective self-report, as well as with objective electroencephalogram measures, recordings of slow eye movements and vigilance tests^{30–38}. Exposure to bright white light, as well as short-wavelength monochromatic light, has been shown to acutely enhance both subjective and objective measures of alertness^{35–41}. Similarly, light stimuli can enhance human performance in terms of psychomotor vigilance and neurocognitive responses such as cognitive throughput, sustained attention

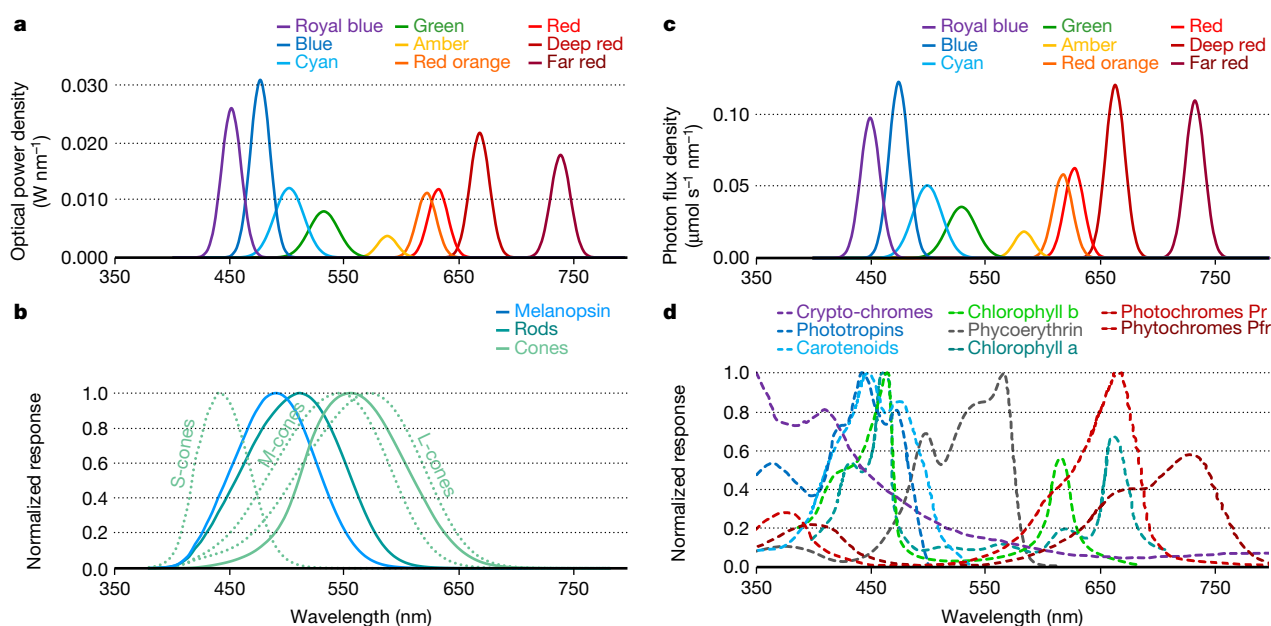


Fig. 2 | LED emission, human response and plant response spectra.

a, Optical power distributions (obtained from ref. ⁹²) plotted against wavelength for state-of-the-art direct-emitting LEDs (driven at 1 W) that span virtually the entire visible spectrum, although currently with reduced efficiency in the green–amber–orange region. **b**, Human photoreceptor action spectra associated both with the primary optical tract (rods and cones, where S, M and L indicate short, medium and long wavelength) and with the intrinsically photosensitive ipRGCs at the start of the retinohypothalamic tract. Action spectra reproduced from supplementary workbook in ref. ⁴. **c**, Photon flux distributions (obtained from ref. ⁹²)

plotted against wavelength for state-of-the-art direct-emitting LEDs (driven at 1 W). Note the shift in the peak heights, compared to those in **a**, as the wavelength increases. Blue photons have higher energy than red photons, according to the Planck relation ($E = hc/\lambda$, where E is the energy, h is the Planck constant, c is the speed of light in vacuum and λ is the wavelength). **d**, Plant action spectra associated with the primary classes of photosensitive molecules in plants. Action spectra for the cryptochromes are not yet well established. Action spectrum for phycoerythrin from ref. ⁹³. All other plant action spectra from ref. ⁹⁴.

and aspects of memory^{34–41}. Not all studies, however, have found a consistent light-induced enhancement of all measures of alertness and neurocognitive responses^{39–41}.

Beyond alertness, a group of studies have shown that light is a regulator of many other aspects of human physiology and behaviour, and has therapeutic capacity in clinical applications such as the treatment of winter depression and selected sleep disorders^{4,5,42,43}. Light therapy has been evaluated in, and is increasingly recommended for, healthy individuals who experience problems related to shift work, intercontinental jet travel and space flight^{4,5,44}. A maxim for optimizing circadian regulation is increased light exposure at the beginning of and during the wake cycle, and decreased light exposure before sleep. However, many open questions remain regarding the detailed physiology of the contributing photoreceptor system and how it influences specific human neurobehavioural responses.

A first open question concerns the nature of the detailed pathways within the melanopsin-based photoreceptor system. This system is both anatomically complex at the level of the retina as well as physiologically complex in terms of regulating neural targets in the brain. All retinal photoreceptors contribute to the regulation of biological and behavioural responses to light, but the relative importance of each photoreceptor is highly labile within and between types of physiological and behavioural responses. Furthermore, the responsiveness of this photoneural system to the wavelength and the intensity of the light is fundamentally context-dependent^{4,33,45–47}.

A second open question considers the interactions between the retino-hypothalamic and primary optical tracts. In one direction, studies with rodents and non-human primates provide compelling evidence that ipRGCs also anatomically project to nuclei of the visual system and physiologically contribute to aspects of visual processing and image detection^{4,29,48–52}. A number of studies with blind and normally sighted human subjects support a role of the melanopsin photoreceptor system in contributing to visual responses in humans^{53–56}. In the other direction, studies on rodents, monkeys and humans clearly show that the

visual rod and cone photoreceptors are anatomically and functionally interconnected with the ipRGCs^{4,33,48–50,56}. Furthermore, there are several subtypes of ipRGCs with diverse connectivity to cells in the inner retina and differing projections to the nuclei in the brain^{49,52,57–59}. We note that even the pupillary light reflex—a seemingly simple interaction between the two retinal tracts—is more complex than it appears. In all species studied, including humans, the pupillary light reflex is a rapid response of the iris to light exposure. Rodent and non-human primate studies demonstrate direct neural projections from the ipRGCs to the nuclei in the midbrain that regulate this reflex^{4,27,29}. Although this reflex is dominated by melanopsin phototransduction in the ipRGCs, the rods and cones also contribute to pupillary responses, particularly at lower light levels^{4,46}. Despite the direct ipRGC projection to the midbrain nuclei, elements of the pupillary light reflex exhibit diurnal rhythms.

A third open question concerns the relationship between the dose of light and physiological regulation in everyday environments. From the light-stimulus side, there are four relevant physical-exposure variables: light intensity, light spectrum, stimulus duration and stimulus timing; all of these variables are fundamentally context-dependent. These variables also depend on elements of ocular and neural physiology involved in light transduction for regulation of the human circadian, neuroendocrine and neurobehavioural systems, including conscious and reflex behaviour of the head and eyes relative to the light source; transmission of light through the ocular media; transduction of light through the iris and/or pupil; wavelength sensitivity of photoreceptors; distribution of photoreceptors; state of photoreceptor adaptation; and neural ability to integrate stimuli temporally and spatially. Controlled laboratory studies have elucidated how each element can contribute to the efficacy of a photic stimulus in eliciting a physiological or behavioural response^{23,60}. There are fewer studies, however, that characterize the efficacy of lighting for physiological regulation under daily living conditions, in which people move freely about an environment that is lit by a combination of different electrical illuminants, window light and outdoor sky lighting.

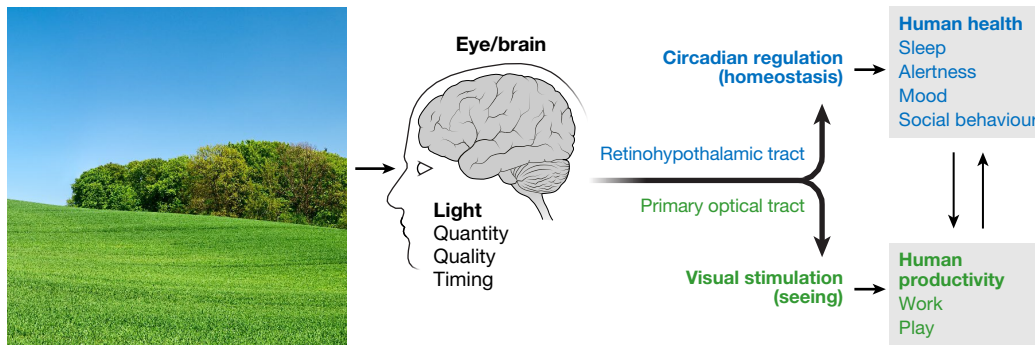


Fig. 3 | The two photoreceptor pathways between the human eye and the brain. The primary optical tract (green text) originates in the retinal rods and cones. Cone photoreceptors in the fovea provide higher-light-level photopic colour vision with a peak sensitivity in the green at a wavelength of approximately 555 nm, the colour of green foliage; rod photoreceptors provide the lower-light-level scotopic black, grey and white vision with a

peak sensitivity at about 498 nm. The retinohypothalamic tract (blue text) originates with ipRGCs, the peak sensitivity of which is at about 480 nm, approximately the colour of the blue sky. This regulates the circadian, neuroendocrine and neurobehavioural systems that ultimately impact human health and productivity. Photograph from iStock/Getty.

A fourth open question asks how to frame our understanding of the positive and negative effects of light^{4,5}. A basic concept of modern medicine is that agents that have the capacity to heal also potentially have the capacity to harm. Dysregulation of circadian physiology by inappropriate light exposure has been linked to several diseases and disorders. For example, epidemiological evidence indicates an association between breast and prostate cancer risk and shift work^{61–63}. Shift work typically involves routine light exposure during the night time that can suppress nocturnal melatonin secretion, disrupt circadian entrainment and interfere with healthy sleep⁶¹. Empirical data with human tumorigenesis supports the epidemiological observations at least for breast cancer⁶⁴. In 2007, such lines of evidence led the World Health Organization to identify long-term shift work as a probable cause of cancer⁶⁵. Similar to the growing information on cancer risk, there is both epidemiological and empirical evidence that circadian disruption and circadian desynchrony contributes to cardiovascular disease, metabolic syndrome, diabetes, obesity and gastrointestinal disorders^{66–71}. It is important to note, however, that there are limitations to the developing science related to the adverse health consequences of night-time light exposure. For example, it is not clear whether circadian disruption due to inappropriate light exposure alone increases the risk of developing cancer, cardiovascular disease or metabolic disorders. Disruption of the human circadian system usually involves disruption of sleep and/or the disruption of normal melatonin rhythms⁶¹. Sleep deprivation and nocturnal melatonin suppression have each been implicated in the potential health consequences associated with shift work and light exposure at night^{61,65,67,69–71}. Despite such uncertainties it is noteworthy that, in 2012, the American Medical Association published a position statement on the adverse health effects of night-time lighting. Specifically, they identified a need for “further multi-disciplinary research on occupational and environmental exposure to light-at-night, the risk of cancer, and effects on various chronic diseases”⁷².

In parallel with research efforts to answer the above open questions, LED lighting is already being actively used in clinical and non-clinical applications. In clinical applications bright-white light therapy, which has been used since the 1980s, has proven to be an effective therapeutic intervention for patients with seasonal affective disorder (known as SAD or winter depression) and its subclinical variant, sSAD^{42,43}. Additional clinical applications have been explored, including light treatment of non-seasonal depression, various sleep disorders, menstrual cycle problems, bulimia nervosa, and fatigue problems associated with senile dementia, chemotherapy and traumatic brain injury^{4,5,42,43}. With the advent of LED lighting technology, therapeutic lighting devices are now being produced in which both broad and narrow bandwidths of light are emitted by LEDs. This advance has enabled light therapy equipment for clinical applications to be produced in

conventionally sized light panels as well as relatively small, portable or wearable devices⁶⁰.

In non-clinical applications, light therapy has been evaluated for healthy individuals who experience problems related to shift work, intercontinental jet travel and space flight^{4,5,42–44}. For example, NASA (National Aeronautics and Space Administration) has used bright white fluorescent-light treatment for improving sleep and fatigue in astronauts since 1990^{73,74}. In 2007, the Phoenix Mars Lander mission provided an opportunity to test the feasibility and efficacy of LED light therapy to synchronize the circadian systems of operational ground personnel to a Mars sol of 24.6 h. Measures of circadian period demonstrated that, as part of a fatigue-management programme, timed therapy with solid-state light enabled 87% of participants to adapt to the Mars sol⁷⁵. More recently, both ground-based and in-flight studies on the International Space Station have been testing LED luminaires for supporting vision and improving circadian entrainment, alertness and sleep in astronauts^{44,76}. Compared with conventional fluorescent light sources, the advantages of LEDs that are of particular relevance to space flight include their reduced weight, power consumption and heat generation; in addition, the LEDs have a tunable spectrum, comprise fewer toxic materials, and have greater resistance to damage and a longer operational life. To date, more than half of the fluorescent light fixtures on the United States’ portion of the International Space Station have been replaced with LED light assemblies that have three preset spectrum and intensity modes: general vision, alerting/circadian phase shifting, and pre-sleep.

Looking forward, there is clearly a growing interest in encouraging the development of LED lighting technologies that have the capacity to improve health, performance and well-being for healthy people in all lighting applications. As discussed above, the effects of light on enhancing alertness as well as improving cognitive and psychomotor responses may lead to advanced, daytime solid-state lighting technologies for schools, workplaces, public buildings, and almost any place that uses electric lighting. Relevant to evening and night-time lighting, a recent study compared standard fluorescent-light fixtures to solid-state sources set to an intensity typical of bedroom lighting in terms of biological and behavioural efficacy. Compared to the fluorescent light, solid-state light evoked a greater secretion of evening melatonin and reduced measures of alertness in healthy subjects, thus physiologically preparing the body for sleep³⁹. Studies like this open the door for the development and application of LED-based lighting systems to benefit individuals in hospitals, care facilities and residential environments. For the general population of relatively healthy individuals, the ubiquity of electric light in the built environment provides an opportunity to tailor the lighting for individually modest health and productivity benefits that are significant when aggregated across the entire population. For healthy, at-risk (elderly, night shift, jet lag) and infirmed or recovering

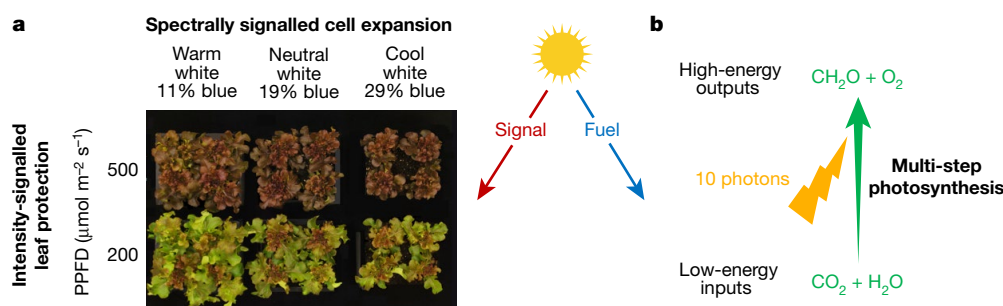


Fig. 4 | Light has both direct and indirect effects on plants. **a**, The indirect effect is a signal that directs plant shape, cell expansion and colour (photomorphology). The lettuce cultivar in this example is Red Salad Bowl. Light intensity triggers the synthesis of anthocyanin (red) pigments in the leaves as a protectant: a photosynthetic photon flux density (PPFD) of 200 (10% of full sunlight) was insufficient, but a PPFD of 500 (25% of full sunlight) triggered anthocyanin synthesis. The wavelength of the light affects leaf expansion: as the fraction of blue light increases (as the colour

temperature shifts from warm to neutral to cool white), leaf expansion decreases. **b**, The direct effect of photons as a fuel for photosynthesis and plant growth (dry mass). Ten photons (400–700 nm) is the approximate theoretical minimum number of photons to fix one molecule of CO_2 into carbohydrate. Photosynthesis includes multiple steps (roughly 23) that turn low-energy inputs into high-energy outputs. The theoretical maximum efficiency of the process is 30%, with each individual step being about 95% efficient ($0.95^{23} \approx 0.3$).

populations, the new features of LED lighting have the potential to improve wellbeing and quality of life, provided that the fundamental questions discussed in this section can begin to be answered in order to guide specific technology and product development of LED light sources.

Lighting for plants

Plants are sentient organisms and have evolved an exquisite sensitivity to ultraviolet, photosynthetic and near-infrared radiation in their environment^{77,78}. The responses of plants to light have fascinated observers since the early days of the scientific method.

Two hundred years ago, primitive light sources and coloured filters were used to characterize the effect of light colour on stem elongation. Eighty years ago, Hoover⁷⁹ found that photosynthesis used a range of wavelengths similar to those of human vision, but with increased sensitivity to blue and red light. Thirty years later, studies by McCree⁸⁰ and Inada⁸¹ refined the early studies of Hoover. Using a monochromator and spectral filters to achieve narrowband radiation, they found that single leaves under monochromatic red light (600–700 nm) had a 25%–35% higher quantum yield than those under blue light (400–500 nm), and a 5%–30% higher quantum yield than those under green light (500–600 nm). It is now known that these classic studies had limitations because they were conducted at low light levels on single leaves. More importantly, the use of monochromatic light did not allow for synergism among wavelengths. LEDs are enabling us to refine these spectral effects on photosynthesis by studying plant communities at higher light levels with synergistic wavelengths.

In the late 1940s, fluorescent lamps made it possible to grow plants without sunlight⁸², but by today's standards the electric conversion efficacy was low. In the late 1970s, high-intensity-discharge (high-pressure sodium and metal halide) lighting made it possible to grow plants at intensities comparable to sunlight, but with a fixed spectral output.

These lighting technologies had a limited ability, however, to separately control the intensity, spectrum, and timing of the delivery of photons. Plants respond to light in more ways than do humans, and utilize more than a dozen photoreceptors to direct their growth (Fig. 2d). They live in communities and respond to light reflected from neighbouring plants. The light that reaches lower leaves is filtered by upper leaves. Plants use photons both as a fuel for photosynthesis and as a signal that directs plant shape and leaf colour (plant development) (Fig. 4). Plant shape includes leaf expansion and radiation capture, both of which increase canopy photosynthesis. LED technology is facilitating a fundamental revolution in research into the photobiology of plants. Promising research directions are discussed below.

First is the effect of the wavelength of light on the morphology of plants. It is known that the fraction of blue light has a powerful

influence on morphology, but the effect varies across species (Fig. 4) and the mechanisms are not yet understood. Ultraviolet radiation has several beneficial effects on plant growth, including increased cuticle thickness, reduced intumescence⁸³ and increased secondary metabolism that leads to improved flavour. Ultraviolet radiation also affects the interaction between plants, fungal pathogens and insects⁸⁴. However, these effects also vary across species. Near-infrared radiation (700–780 nm) has a powerful effect on stem elongation and the rate of leaf expansion, but although the primary receptor (phytochrome) is well characterized, significant interactions with other wavelengths are now being discovered.

A second question concerns the value of green light. Because chlorophyll absorbs green light only minimally, many researchers concluded that it was of low value for photosynthesis⁸⁵. However, because it penetrates deeper into leaves and plant canopies, more recent studies have shown that green photons have a similar value in photosynthesis to those of other colours^{77,78}.

Third, the interaction between the spectrum and the intensity of light (quality and quantity) is also of interest. Early studies with LEDs were performed at a photosynthetic photon flux density (PPFD) of less than 10% of full sunlight ($200 \mu\text{mol m}^{-2} \text{s}^{-1}$), but morphological effects in low light can be reversed at higher light intensities. Blue photons, for example, interact with total PPFD to determine cell expansion (Fig. 4), elongation of the petiole, stem and leaf, and—indirectly via these morphology changes—radiation capture⁷⁸.

A fourth research direction investigates the interactions between spectra, intensity and time of exposure (time of day and stage of plant growth). The timing of the delivery of photons can help to understand the photobiological mechanisms and sites of perception. LEDs make it possible to pulse ultraviolet radiation, which can improve flavour and minimize the detrimental effects of ultraviolet radiation on DNA⁸⁶.

Finally it is important to scale from monochromatic short-term measurements on single leaves to long-term performance of plant communities. This is particularly challenging and important because plants adapt to changes in radiation by synthesizing new pigments. Changing the quality of the light also alters the shape of the plant, and multiple wavelengths interact synergistically, so defining the effects of the quality of light on photosynthesis of the whole plant and the plant community is a complex enterprise.

Even as these research directions are being pursued, LEDs are being increasingly used in a practical horticultural context. For the first time in history, photons can be precisely applied in order to grow food. Assuming the economics outlined in Box 1, the value of summer sunlight in the mid-latitudes is US\$700,000 per hectare over a 100-day growing season, and about US\$70,000 per hectare during the darkest 60 days of the year⁸⁷. Although these costs are high, the value of fresh leafy greens that can be produced with supplementary LED lighting

Box 1

Economics of indoor agriculture

Lighting for plants presents unique challenges, because plants require 30–100 times higher light intensities than do humans. Efficient LED lighting reduces the photon cost and enables indoor agriculture for high-value crops. To demonstrate this, we calculate the efficacy of production (E_{dry} , in grams of dry mass produced per mole of photons) using the equation

$$E_{\text{dry}} = F_a \times QY \times CUE \times HI \times k$$

where F_a is the fraction of photons absorbed; QY is the quantum yield (moles of carbon fixed per mole of photons absorbed); CUE is the carbon use efficiency (moles of carbon incorporated into plant biomass per mole of carbon fixed); HI is the harvest index (moles of carbon in edible product per mole of carbon in plant biomass); and $k = 30$, a constant that represents the mass of CH_2O (carbohydrate) in grams per mole of carbon in edible product.

Values of the five parameters for five types of crop and the resulting production efficacies are shown in the table. We also show a calculation assuming the highest achievable values for each parameter ('potential efficacy'), which requires CO_2 levels to be increased to four times ambient levels in order to minimize photorespiration. Production efficacies of leafy microgreens approach this potential efficacy, but for other crops the efficacies are lower. Lettuce benefits from higher light levels (15% of full sunlight), which reduces QY from 0.08 to 0.07, and lettuce plants are typically spaced farther apart during early growth, which reduces radiation capture (F_a) from 0.95 to 0.65. Tomatoes benefit from even higher light levels (at least 25% of full sunlight), which reduces QY further, to 0.05, and the stems, roots and leaves of tomato plants are not edible, reducing HI to about 0.6. Vegetables such as broccoli and strawberries, and staple crops such as rice and wheat, have even lower HI .

The photon cost per dry mass is the cost per mole of photons (assumed to be US\$0.01 mol^{-1} for the most efficient LED fixtures, which have an electricity cost of US\$0.10 kWh^{-1}) divided by E_{dry} . Although the cost per mole of photons is the same for all crops, different crops have different E_{dry} , so more photons are required to create the same dry mass for some crops. The photon cost per dry mass thus varies, from a low of US\$7.5 $\text{kg}_{\text{dry}}^{-1}$ for leafy microgreens to US\$41 $\text{kg}_{\text{dry}}^{-1}$ for general vegetables, rice and wheat.

The market prices of the crops vary with their percentage water content. The fresh market prices vary widely, from US\$35 $\text{kg}_{\text{fresh}}^{-1}$ for leafy microgreens to US\$0.40 $\text{kg}_{\text{fresh}}^{-1}$ for rice and wheat. The dry market prices vary even more, because of the vast differences in the water content of the crops, from US\$700 $\text{kg}_{\text{dry}}^{-1}$ for leafy microgreens to US\$0.40 $\text{kg}_{\text{dry}}^{-1}$ for rice and wheat.

The final column of the table shows the ratio of the photon cost per dry mass to the dry market price, which is determined from the fresh market price and the water fraction. Leafy greens, which have the highest E_{dry} (and therefore the lowest photon cost per dry mass), also have the highest market prices. The effective photon cost thus increases rapidly with more complex crops (such as rice and wheat). The economics of simple leafy crops delivered fresh can be quite favourable. The effective cost of photons greatly exceeds the value of agronomic crops that are delivered dry and even exceeds the retail value of potatoes. Even if LEDs were 100% efficient, it would not be cost-effective to grow our staple agronomic crops with electric light. Thus, electric light input is a small cost for microgreens, a high cost for general vegetables and an unacceptable cost for staple crops. Because leafy greens are perishable and the fresh product has a high retail price, indoor farming is dominated by leafy greens.

Crop type	F_a	QY	CUE	HI	k	E_{dry}	Photon cost per dry mass (US\$ $\text{kg}_{\text{dry}}^{-1}$)	Fresh market price (US\$ $\text{kg}_{\text{fresh}}^{-1}$)	Water content (%)	Dry market price (US\$ $\text{kg}_{\text{dry}}^{-1}$)	Photon cost (% of dry market price)
Potential efficacy	0.95	0.08	0.65	0.9	30	1.33	8				
Leafy microgreens	0.95	0.08	0.65	0.9	30	1.33	8	35	95	700	1
Lettuce	0.65	0.07	0.65	0.9	30	0.80	13	12	95	240	5
Tomatoes	0.60	0.05	0.65	0.6	30	0.35	29	8	95	160	18
General vegetables	0.50	0.05	0.65	0.5	30	0.24	42	4	90	40	103
Rice or wheat	0.50	0.05	0.65	0.5	30	0.24	42	0.40	3	0.40	10,000

is even higher. Seasonal combinations of sunlight and electric lights have the potential to markedly expand the range of local, year-round production of fresh greens.

An additional benefit of LED-enabled horticulture in a closed system is that water can be recycled by condensing water vapour in the air-conditioning system and returning it to the root zone. If, in the future, water were to become more expensive than energy, this would be a valuable advantage. A closed system could also reduce the need to apply pharmaceuticals owing to the limited access to pests, although lush growth in an optimal environment makes plants more susceptible to fungal pathogens such as *Pythium*. Controlled, area-intensive farming has attracted great commercial interest because it can move farming closer to urban population centres, and increase the quality of the goods by minimizing the transport time of perishable produce⁸⁸.

LEDs may also facilitate a co-evolution of plant genetics and plant environment. Plants can be engineered to better utilize the unique environment, which may, in turn, create new requirements for LEDs. This synergism between genetics and environment is the underlying

reason for the marked increase in productivity of the world's agricultural system over the past 100 years. The opportunity to expand genetic potential with light has led to a new class of plants, which have been referred to as environmentally modified organisms or EMOs⁸⁹. For example, plants have evolved self-protection mechanisms: mostly against insects and diseases, but also to cope with variations in light intensity and temperature. Exposure to ultraviolet radiation triggers the synthesis of ultraviolet-blocking pigments that prevent high-energy photons from inducing damage via the generation of reactive oxygen species⁹⁰. If pests and ultraviolet radiation are eradicated through the use of electric lighting in a controlled environment, the need for these defence mechanisms can be reduced or even eliminated, and the efficiency of crop production might be increased.

The potential for breeding new cultivars gives rise to a great increase in scientific and technological possibilities for engineering plants for controlled environments. The tangible benefit is that health-promoting fresh produce becomes more accessible in all regions of the world in all months of the year.

Conclusions

The development of LED lighting was motivated by the promise of significant energy savings. These savings are now coming to pass. Unlike other energy-saving technologies, LED technology does not require any performance compromise, but rather improves performance while also offering new levels of control and value. Now we are entering a new world of lighting, one that includes both applications that go beyond basic illumination and value propositions that have not previously been associated with lighting.

Two primary examples are lighting for human health and productivity, and lighting for plants. The benefit of LED lighting here is not the saving of energy, although both applications will certainly benefit from improved efficiency of the light sources. The benefits here are more profound: improving human health and productivity through our emerging understanding of the physiological lighting requirements of humans, including a potential decrease in some forms of cancer and other clinical disorders; and diversifying, improving, and localizing food production in controlled environments. These nascent applications have revealed how little is known about light and physiological responses, and LEDs are providing the tools to help us better understand how humans and plants respond to light.

Even in basic lighting applications, the new levels of control offered by LED lighting have raised questions about our understanding of lighting science. Questions regarding spectral content, standard lighting levels, colour perception and preference, glare, flicker, and their impacts on visual performance are being raised. Although these issues are not fundamental shortcomings of LED lighting technology, they highlight difficulties in designing, specifying and deploying LED products when there are so many new levels of control that bring new questions. For every use of lighting, the technological possibilities of LED lighting currently outstrip our understanding of how best to use the light for the application. This situation requires research in all fields of lighting application science, concurrent with ongoing research into the underlying technology in order to achieve fully optimized lighting systems that enable the full promise of LED lighting. The promised benefits of the new world of lighting can be achieved with no obvious, fundamental downside and include vast energy savings and associated atmospheric carbon reductions, improved human health, healthier and more localized food production, and the reduced ecological impacts of light at night⁹¹. With these prospects at hand, the new world of LED lighting may be as profound a revolution as the transition from gas lighting to electric incandescent lighting that occurred a century ago.

Received: 26 January 2017; Accepted: 15 October 2018;

Published online 21 November 2018.

- Parker, A. *In the Blink of an Eye: How Vision Sparked the Big Bang of Evolution* (Basic Books, New York, 2003).
- Gerkema, M. P., Davies, W. I., Foster, R. G., Menaker, M. & Hut, R. A. The nocturnal bottleneck and the evolution of activity patterns in mammals. *Proc. R. Soc. Lond. B* **280**, 20130508 (2013).
- Gregory, R. L. *Eye and Brain: The Psychology of Seeing* (Princeton Univ. Press, Princeton, 2015).
- Lucas, R. J. et al. Measuring and using light in the melanopsin age. *Trends Neurosci.* **37**, 1–9 (2014).
- A summary of the neurophysiology of the melanopsin ipRGC sensory pathway and of the implications for the measurement, production and application of light (includes a free measurement tool to calculate the photoreceptive inputs for circadian, neuroendocrine and neurobehavioral responses).**
- Figueiro, M. G., Brainard, G. C., Lockley, S. W., Revell, V. L. & White, R. *Light and Human Health: An Overview of the Impact of Optical Radiation on Visual, Circadian, Neuroendocrine and Neurobehavioral Responses*. Technical Memorandum IES TM-18-08 (Illuminating Engineering Society, 2008).
- Bowers, B. & Anastas, P. *Lengthening the Day: A History of Lighting Technology* (Oxford Univ. Press, Oxford, 1998).
- Boyce, P. R. *Human factors in lighting* (CRC Press, Boca Raton, 2014).
- Schivelbusch, W. *Disenchanted Night: The Industrialization of Light in the Nineteenth Century* (Univ. California Press, Berkeley, 1995).
- Steinmetz, C. P. *Radiation, Light and Illumination: A Series of Engineering Lectures Delivered at Union College* (McGraw-Hill, New York, 1918).
- Tsao, J. Y., Han, J., Haitz, R. H. & Pattison, P. M. The blue LED Nobel prize: historical context, current scientific understanding, human benefit. *Ann. Phys.* **527**, A53–A61 (2015).
- A succinct discussion of the Nobel-prize-winning breakthroughs that led to blue-LED and LED lighting and of the context of these breakthroughs in the history of semiconductor science and technology.**

- US DOE SSL Program. *Solid-State Lighting R&D Plan*. https://www.energy.gov/sites/prod/files/2018/09/f56/ssl_rd-plan_jun2016.pdf (2016).
- Krames, M. R. et al. Status and future of high-power light-emitting diodes for solid-state lighting. *J. Disp. Technol.* **3**, 160–175 (2007).
- Tsao, J. Y. & Waide, P. The world's appetite for light: Empirical data and trends spanning three centuries and six continents. *Leukos* **6**, 259–281 (2010).
- US DOE SSL Program. *Energy Savings Forecast of Solid-State Lighting in General Illumination Applications*. https://energy.gov/sites/prod/files/2016/09/f33/energysavingsforecast16_2.pdf (2016).
- Schubert, E. F. & Kim, J. K. Solid-state light sources getting smart. *Science* **308**, 1274–1278 (2005).
- The first paper to discuss the potential of solid-state lighting to be 'smart', in the sense of being able not only to provide energy savings but also to adjust to the specific environments and requirements of a wide range of applications.**
- Tsao, J. Y. et al. Toward smart and ultra-efficient solid-state lighting. *Adv. Opt. Mater.* **2**, 809–836 (2014).
- A comprehensive review of the state of solid-state lighting in terms of its ultimate potential to be both 'smart' and ultra-efficient.**
- Watson, B. From light to bright: San Diego is building the world's largest municipal Internet of Things. *GE Reports* <https://www.ge.com/reports/light-bright-san-diego-leads-way-future-smart-cities/> (2017).
- Tsao, J. Y., Schubert, E. F., Fouquet, R. & Lave, M. The electrification of energy: long-term trends and opportunities. *MRS Energy Sustain.* **5**, E7 (2018).
- US DOE SSL Program. *Solid-State Lighting 2017 Suggested Research Topics Supplement: Technology and Market Context*. https://energy.gov/sites/prod/files/2017/09/f37/ssl_supplement_suggested-topics_sep2017_0.pdf (2017).
- Dijk, D. J. & von Schantz, M. Timing and consolidation of human sleep, wakefulness, and performance by a symphony of oscillators. *J. Biol. Rhythms* **20**, 279–290 (2005).
- Thapan, K., Arendt, J. & Skene, D. J. An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *J. Physiol.* **535**, 261–267 (2001).
- Brainard, G. C. et al. Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**, 6405–6412 (2001).
- Brainard, G. C. & Hanifin, J. P. Photons, clocks, and consciousness. *J. Biol. Rhythms* **20**, 314–325 (2005).
- Provencio, I. et al. A novel human opsin in the inner retina. *J. Neurosci.* **20**, 600–605 (2000).
- Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nat. Neurosci.* **4**, 1165 (2001).
- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: an opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
- A landmark paper that details the discovery of melanopsin, ultimately leading to melanopsin being identified as a functional photopigment in the retinas of mammals, including humans, with roles in the regulation by light of circadian, neuroendocrine, neurobehavioral and visual responses.**
- Hannibal, J. et al. Central projections of intrinsically photosensitive retinal ganglion cells in the macaque monkey. *J. Comp. Neurol.* **522**, 2231–2248 (2014).
- Cajochen, C., Khalsa, S. B. S., Wyatt, J. K., Czeisler, C. A. & Dijk, D. J. EEG and ocular correlates of circadian melatonin phase and human performance decrements during sleep loss. *Am. J. Physiol.* **277**, R640–R649 (1999).
- Cajochen, C., Zeitzer, J. M., Czeisler, C. A. & Dijk, D. J. Dose-response relationship for light intensity and ocular and electroencephalographic correlates of human alertness. *Behav. Brain Res.* **115**, 75–83 (2000).
- Cajochen, C. et al. High sensitivity of human melatonin, alertness, thermoregulation, and heart rate to short wavelength light. *J. Clin. Endocrinol. Metab.* **90**, 1311–1316 (2005).
- Lockley, S. W. et al. Short-wavelength sensitivity for the direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans. *Sleep* **29**, 161–168 (2006).
- Wright, K. P. Jr, Badia, P., Myers, B. L. & Plenzer, S. C. Combination of bright light and caffeine as a countermeasure for impaired alertness and performance during extended sleep deprivation. *J. Sleep Res.* **6**, 26–35 (1997).
- Chang, A. M., Scheer, F. A., Czeisler, C. A. & Aeschbach, D. Direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans depend on prior light history. *Sleep* **36**, 1239–1246 (2013).
- Rüger, M., Gordijn, M. C., Beersma, D. G., de Vries, B. & Daan, S. Weak relationships between suppression of melatonin and suppression of sleepiness/fatigue in response to light exposure. *J. Sleep Res.* **14**, 221–227 (2005).
- Phipps-Nelson, J., Redman, J. R., Dijk, D. J. & Rajaratnam, S. M. Daytime exposure to bright light, as compared to dim light, decreases sleepiness and improves psychomotor vigilance performance. *Sleep* **26**, 695–700 (2003).
- Cajochen, C. et al. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *J. Appl. Physiol.* **110**, 1432–1438 (2011).
- Rahman, S. A., St Hilaire, M. A. & Lockley, S. W. The effects of spectral tuning of evening ambient light on melatonin suppression, alertness and sleep. *Physiol. Behav.* **177**, 221–229 (2017).
- Segal, A. Y., Sletten, T. L., Flynn-Evans, E. E., Lockley, S. W. & Rajaratnam, S. M. Daytime exposure to short- and medium-wavelength light did not improve alertness and neurobehavioral performance. *J. Biol. Rhythms* **31**, 470–482 (2016).

41. Sletten, T. L. et al. Randomised controlled trial of the efficacy of a blue-enriched light intervention to improve alertness and performance in night shift workers. *Occup. Environ. Med.* **74**, 792–801 (2017).
42. Lam, R. W. & Tam, E. M. *A Clinician's Guide to Using Light Therapy* (Cambridge Univ. Press, New York, 2009).
43. Wirz-Justice, A., Benedetti, F., Terman, M. & Basel, S. Chronotherapeutics for affective disorders: a clinician's manual for light and wake therapy. *Ann. Clin. Psychiatry* **22**, 67 (2010).
44. Brainard, G. C., Barger, L. K., Soler, R. R. & Hanifin, J. P. The development of lighting countermeasures for sleep disruption and circadian misalignment during spaceflight. *Curr. Opin. Pulm. Med.* **22**, 535–544 (2016).
45. Gooley, J. J., et al. Spectral responses of the human circadian system depend on the irradiance and duration of exposure to light. *Sci. Transl. Med.* **2**, 31ra33 (2010).
46. Lall, G. S. et al. Distinct contributions of rod, cone, and melanopsin photoreceptors to encoding irradiance. *Neuron* **66**, 417–428 (2010).
47. Altimus, C. M. et al. Rod photoreceptors drive circadian photoentrainment across a wide range of light intensities. *Nat. Neurosci.* **13**, 1107–1112 (2010).
48. Dacey, D. M. et al. Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature* **433**, 749–754 (2005).
49. Ecker, J. L. et al. Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron* **67**, 49–60 (2010).
50. Brown, T. M., Wynne, J., Piggins, H. D. & Lucas, R. J. Multiple hypothalamic cell populations encoding distinct visual information. *J. Physiol.* **589**, 1173–1194 (2011).
51. Brown, T. M. et al. Melanopsin-based brightness discrimination in mice and humans. *Curr. Biol.* **22**, 1134–1141 (2012).
52. Estevez, M. E. et al. Form and function of the M4 cell, an intrinsically photosensitive retinal ganglion cell type contributing to geniculocortical vision. *J. Neurosci.* **32**, 13608–13620 (2012).
53. Zaidi, F. H. et al. Short-wavelength light sensitivity of circadian, pupillary, and visual awareness in humans lacking an outer retina. *Curr. Biol.* **17**, 2122–2128 (2007).
54. Zele, A. J., Feigl, B., Adhikari, P., Maynard, M. L. & Cao, D. Melanopsin photoreceptor contributes to human visual detection, temporal and colour processing. *Sci. Rep.* **8**, 3842 (2018).
55. Horiguchi, H., Winawer, J., Dougherty, R. F. & Wandell, B. A. Human trichromacy revisited. *Proc. Natl Acad. Sci. USA* **110**, E260–E269 (2013).
56. Spitschan, M., Datta, R., Stern, A. M., Brainard, D. H. & Aguirre, G. K. Human visual cortex responses to rapid cone and melanopsin-directed flicker. *J. Neurosci.* **36**, 1471–1482 (2016).
57. Zhao, X., Stafford, B. K., Godin, A. L., King, W. M. & Wong, K. Y. Photoresponse diversity among the five types of intrinsically photosensitive retinal ganglion cells. *J. Physiol.* **592**, 1619–1636 (2014).
58. Prigge, C. L. et al. M1 ipRGCs influence visual function through retrograde signaling in the retina. *J. Neurosci.* **36**, 7184–7197 (2016).
59. Hannibal, J., Christiansen, A. T., Heegaard, S., Fahrenkrug, J. & Kilgaard, J. F. Melanopsin expressing human retinal ganglion cells: Subtypes, distribution, and intraretinal connectivity. *J. Comp. Neurol.* **525**, 1934–1961 (2017).
60. Brainard, G. C. & Hanifin, J. P. *Handbook of Advanced Lighting Technology* (Springer, Berlin, 2017).
61. Stevens, R. G., Brainard, G. C., Blask, D. E., Lockley, S. W. & Motta, M. E. Breast cancer and circadian disruption from electric lighting in the modern world. *CA Cancer J. Clin.* **64**, 207–218 (2014).
- A summary of the empirical and epidemiological evidence relating to the potential health consequences of inappropriate night-time light exposure disrupting human circadian physiology, melatonin production and sleep.**
62. Rao, D., Yu, H., Bai, Y., Zheng, X. & Xie, L. Does night-shift work increase the risk of prostate cancer? A systematic review and meta-analysis. *Oncotargets Ther.* **8**, 2817–2826 (2015).
63. James, P. et al. Outdoor light at night and breast cancer incidence in the nurses' health study II. *Environ. Health Perspect.* **125**, 087010 (2017).
64. Blask, D. E. et al. Melatonin-depleted blood from premenopausal women exposed to light at night stimulates growth of human breast cancer xenografts in nude rats. *Cancer Res.* **65**, 11174–11184 (2005).
65. World Health Organization, International Agency for Research on Cancer. Shiftwork. *IARC Monogr. Eval. Carcinog. Risks Hum.* **98**, 561 (2010).
66. Scheer, F. A., Hilton, M. F., Mantzoros, C. S. & Shea, S. A. Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc. Natl Acad. Sci. USA* **106**, 4453–4458 (2009).
67. Buxton, O. M., et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. *Sci. Transl. Med.* **4**, 129ra43 (2012).
68. Morris, C. J. et al. Endogenous circadian system and circadian misalignment impact glucose tolerance via separate mechanisms in humans. *Proc. Natl Acad. Sci. USA* **112**, E2225–E2234 (2015).
69. Morris, C. J., Purvis, T. E., Mistretta, J. & Scheer, F. A. Effects of the internal circadian system and circadian misalignment on glucose tolerance in chronic shift workers. *J. Clin. Endocrinol. Metab.* **101**, 1066–1074 (2016).
70. Leproult, R. & Van Cauter, E. in *Pediatric Neuroendocrinology* Vol. 17 (eds Loche, S. et al.) 11–21 (Karger, Basel, 2010).
71. Leproult, R., Holmbäck, U. & Van Cauter, E. Circadian misalignment augments markers of insulin resistance and inflammation, independently of sleep loss. *Diabetes* **63**, 1860–1869 (2014).
72. Blask, D. et al. *Light Pollution: Adverse Health Effects of Nighttime Lighting*. CSAPH Report 4-A-12 (American Medical Association, 2012).
73. Czeisler, C. A., Chiasera, A. J. & Duffy, J. F. Research on sleep, circadian rhythms and aging: applications to manned spaceflight. *Exp. Gerontol.* **26**, 217–232 (1991).
74. Stewart, K. T., Hayes, B. C. & Eastman, C. I. Light treatment for NASA shiftworkers. *Chronobiol. Int.* **12**, 141–151 (1995).
75. Barger, L. K. et al. Learning to live on a Mars day: fatigue countermeasures during the Phoenix Mars Lander mission. *Sleep* **35**, 1423–1435 (2012).
76. Brainard, G. C. et al. Solid-state lighting for the International Space Station: tests of visual performance and melatonin regulation. *Acta Astronaut.* **92**, 21–28 (2013).
77. Bugbee, B. Toward an optimal spectral quality for plant growth and development: the importance of radiation capture. *Acta Hortic.* **1134**, 1–12 (2016).
- A comprehensive review of the current research on spectral effects on photosynthesis and plant morphology.**
78. Snowden, M. C., Cope, K. R. & Bugbee, B. Sensitivity of seven diverse species to blue and green light: interactions with photon flux. *PLoS ONE* **11**, e0163121 (2016).
- A review of interactions between photon flux, photosynthesis and plant morphology.**
79. Hoover, W. H. The dependence of carbon dioxide assimilation in a higher plant on wavelength of radiation. *Smithson. Misc. Collect.* **95**, 1–13 (1937).
80. McCree, K. J. The action spectrum, absorbance and quantum yield of photosynthesis in crop plants. *Agric. Meteorol.* **9**, 191–216 (1971).
81. Inada, K. Action spectra for photosynthesis in higher plants. *Plant Cell Physiol.* **17**, 355–365 (1976).
82. Downs, R. J. *Controlled Environments of Plant Research* (Columbia Univ. Press, New York, 1975).
83. Kubota, C., Eguchi, T. & Kroggel, M. UV-B radiation dose requirement for suppressing intumescence injury on tomato plants. *Sci. Hortic.* **226**, 366–371 (2017).
84. Raviv, M. & Antignus, Y. UV radiation effects on pathogens and insect pests of greenhouse-grown crops. *Photochem. Photobiol.* **79**, 219–226 (2004).
85. Went, F. W. *The Experimental Control of Plant Growth* (Chronica Botanica, New York, 1957).
- A classic book on the early days of photobiology.**
86. Höll, J. et al. Impact of pulsed UV-B stress exposure on plant performance: How recovery periods stimulate secondary metabolism while reducing adaptive growth attenuation. *Plant Cell Environ.* (2018).
87. Bugbee, B. in *Light Emitting Diodes for Agriculture* (ed. Gupta, D.) 81–99 (Springer, Singapore, 2017).
88. Kozai, T., Fujiwara, K. & Runkle, E. S. *LED Lighting for Urban Agriculture* (Springer, Singapore, 2016).
89. Carvalho, S. D. & Folta, K. M. Environmentally modified organisms – expanding genetic potential with light. *Crit. Rev. Plant Sci.* **33**, 486–508 (2014).
- A comprehensive review of spectral effects on plant growth for 20 major crops.**
90. Murchie, E. H. & Niyogi, K. K. Manipulation of photoprotection to improve plant photosynthesis. *Plant Physiol.* **155**, 86–92 (2011).
91. Zielinska-Dabkowska, K. M. Make lighting healthier. *Nature* **553**, 274–276 (2018).
92. *LUXEON Rebel Color Line Product Datasheet DS68*. <https://www.lumileds.com/uploads/265/DS68-pdf> (Lumileds Holding B.V., 2017).
93. Roederer, M. Conjugation of monoclonal antibodies. <http://www.drmmr.com/abcon/> (2004).
94. Pattison, P. M., Tsao, J. Y. & Krames, M. R. Light-emitting diode technology status and directions: opportunities for horticultural lighting. *Acta Hortic.* **1134**, 413–426 (2016).

Acknowledgements P.M.P. and J.Y.T. acknowledge support from the Department of Energy through its Office of Energy Efficiency and Renewable Energy's Solid-State Lighting Program under contract DE-FE0025912. J.Y.T. acknowledges support from Sandia National Laboratories, a multi-program laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The work of S.S.L.S. Inc. is carried out on behalf of the US Department of Energy SSL Program, Washington, DC. G.C.B. acknowledges J. Hanifin for discussions, editorial review and referencing; F. Scheer, K. Roelcklein and R. Lucas for insights on portions of the text; and B. Warfield for help with the design of Fig. 3. G.C.B. was supported, in part, by NASA grants NNX15AC14, NNX08AD66A and NNX09AM68G; NSF grant EEC-0812056; DOE grant DE-EE0008207; The Institute for Integrative Health; and the Philadelphia Section of the Illuminating Engineering Society. B.B. is indebted to 35 years of discussions with colleagues from around the world including T. Volk, M. van Iersel, M. Blonquist, J. Frantz, R. Heins, R. Wheeler and C. Mitchell. B.B. also acknowledges support from NASA grant NNX17AJ31G, the USDA Specialty Crop Research Initiative and the Utah Agricultural Experiment Station. Any opinions, findings and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the authors' funding agencies, including the US Department of Energy, Sandia National Laboratories and NASA.

Reviewer information Nature thanks J.-H. You, J. Wargent and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions P.M.P. outlined the overall article; J.Y.T., G.C.B., B.B. and P.M.P. wrote first drafts of the introductory, 'Lighting for human health and productivity', 'Lighting for plants' and 'Conclusions' sections, respectively; P.M.P., J.Y.T., G.C.B. and B.B. subsequently reviewed and edited all sections.

Competing interests The authors declare no competing interests.

Additional information

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to P.M.P.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Improved reference genome of *Aedes aegypti* informs arbovirus vector control

Benjamin J. Matthews^{1,2,3,49*}, Olga Dudchenko^{4,5,6,7,49}, Sarah B. Kingan^{8,49}, Sergey Koren⁹, Igor Antoshechkin¹⁰, Jacob E. Crawford¹¹, William J. Glassford¹², Margaret Herre^{1,3}, Seth N. Redmond^{13,14}, Noah H. Rose^{15,16}, Gareth D. Weedall^{17,18}, Yang Wu^{19,20,21}, Sanjit S. Batra^{4,5,6}, Carlos A. Brito-Sierra^{22,23}, Steven D. Buckingham²⁴, Corey L. Campbell²⁵, Saki Chan²⁶, Eric Cox²⁷, Benjamin R. Evans²⁸, Thanyalak Fansiri²⁹, Igor Filipović³⁰, Albin Fontaine^{31,32,33,34}, Andrea Gloria-Soria^{28,35}, Richard Hall⁸, Vinita S. Joardar²⁷, Andrew K. Jones³⁶, Raissa G. G. Kay³⁷, Vamsi K. Kodali²⁷, Joyce Lee²⁶, Gareth J. Lycett¹⁷, Sara N. Mitchell¹¹, Jill Muehling⁸, Michael R. Murphy²⁷, Arina D. Omer^{4,5,6}, Frederick A. Partridge²⁴, Paul Peluso⁸, Aviva Presser Aiden^{4,5,38,39}, Vidya Ramasamy³⁶, Gordana Rašić³⁰, Sourav Roy⁴⁰, Karla Saavedra-Rodriguez²⁵, Shruti Sharan^{22,23}, Atashi Sharma^{21,41}, Melissa Laird Smith⁸, Joe Turner⁴², Allison M. Weakley¹¹, Zhilei Zhao^{15,16}, Omar S. Akbari^{43,44}, William C. Black IV²⁵, Han Cao²⁶, Alistair C. Darby⁴², Catherine A. Hill^{22,23}, J. Spencer Johnston⁴⁵, Terence D. Murphy²⁷, Alexander S. Raikhel⁴⁰, David B. Sattelle²⁴, Igor V. Sharakhov^{21,41,46}, Bradley J. White¹¹, Li Zhao⁴⁷, Erez Lieberman Aiden^{4,5,6,7,13}, Richard S. Mann¹², Louis Lambrechts^{31,33}, Jeffrey R. Powell²⁸, Maria V. Sharakhova^{21,41,46}, Zhijian Tu^{20,21}, Hugh M. Robertson⁴⁸, Carolyn S. McBride^{15,16}, Alex R. Hastie²⁶, Jonas Korlach⁸, Daniel E. Neafsey^{13,14}, Adam M. Phillippy⁹ & Leslie B. Vosshall^{1,2,3}

Female *Aedes aegypti* mosquitoes infect more than 400 million people each year with dangerous viral pathogens including dengue, yellow fever, Zika and chikungunya. Progress in understanding the biology of mosquitoes and developing the tools to fight them has been slowed by the lack of a high-quality genome assembly. Here we combine diverse technologies to produce the markedly improved, fully re-annotated AaegL5 genome assembly, and demonstrate how it accelerates mosquito science. We anchored physical and cytogenetic maps, doubled the number of known chemosensory ionotropic receptors that guide mosquitoes to human hosts and egg-laying sites, provided further insight into the size and composition of the sex-determining M locus, and revealed copy-number variation among glutathione S-transferase genes that are important for insecticide resistance. Using high-resolution quantitative trait locus and population genomic analyses, we mapped new candidates for dengue vector competence and insecticide resistance. AaegL5 will catalyse new biological insights and intervention strategies to fight this deadly disease vector.

An accurate and complete genome assembly is required to understand the unique aspects of mosquito biology and to develop control strategies to reduce their capacity to spread pathogens¹. The *Ae. aegypti* genome is large (approximately 1.25 Gb) and highly repetitive, and a 2007 genome project (AaegL3)² was unable to produce a contiguous genome fully anchored to a physical chromosome map³ (Fig. 1a). A more recent assembly, AaegL4⁴, produced chromosome-length scaffolds that made it possible to detect larger-scale syntenic genomic

regions in other species but suffered from short contigs (contig N50, 84 kb, meaning that half of the assembly is found on contigs >84 kb) and a correspondingly large number of gaps (31,018; Fig. 1b). Taking advantage of rapid advances in sequencing and assembly technology in the last decade, we used long-read Pacific Biosciences sequencing and Hi-C (a high-throughput sequencing method based on chromosome conformation capture) scaffolding to produce a new reference genome (AaegL5) that is highly contiguous, with a decrease of

¹Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, NY, USA. ²Howard Hughes Medical Institute, New York, NY, USA. ³Kavli Neural Systems Institute, New York, NY, USA. ⁴The Center for Genome Architecture, Baylor College of Medicine, Houston, TX, USA. ⁵Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ⁶Department of Computer Science, Rice University, Houston, TX, USA. ⁷Center for Theoretical and Biological Physics, Rice University, Houston, TX, USA. ⁸Pacific Biosciences, Menlo Park, CA, USA. ⁹National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ¹⁰Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. ¹¹Verily Life Sciences, South San Francisco, CA, USA. ¹²Mortimer B. Zuckerman Mind Brain Behavior Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. ¹³Broad Institute of MIT and Harvard, Cambridge, MA, USA. ¹⁴Department of Immunology and Infectious Disease, Harvard T. H. Chan School of Public Health, Boston, MA, USA. ¹⁵Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. ¹⁶Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA. ¹⁷Vector Biology Department, Liverpool School of Tropical Medicine, Liverpool, UK. ¹⁸Liverpool John Moores University, Liverpool, UK. ¹⁹Department of Pathogen Biology, School of Public Health, Southern Medical University, Guangzhou, China. ²⁰Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA. ²¹Frail Life Science Institute, Virginia Tech, Blacksburg, VA, USA. ²²Department of Entomology, Purdue University, West Lafayette, IN, USA. ²³Purdue Institute for Inflammation, Immunology and Infectious Disease, Purdue University, West Lafayette, IN, USA. ²⁴Centre for Respiratory Biology, UCL Respiratory, University College London, London, UK. ²⁵Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, USA. ²⁶Bionano Genomics, San Diego, CA, USA. ²⁷National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. ²⁸Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. ²⁹Vector Biology and Control Section, Department of Entomology, Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand. ³⁰Mosquito Control Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ³¹Insect-Virus Interactions Group, Department of Genomes and Genetics, Institut Pasteur, Paris, France. ³²Unité de Parasitologie et Entomologie, Département des Maladies Infectieuses, Institut de Recherche Biomédicale des Armées, Marseille, France. ³³Centre National de la Recherche Scientifique, Unité Mixte de Recherche 2000, Paris, France. ³⁴Aix Marseille Université, IRD, AP-HM, SSA, UMR Vecteurs – Infections Tropicales et Méditerranéennes (VITROME), IHU – Méditerranée Infection, Marseille, France. ³⁵The Connecticut Agricultural Experiment Station, New Haven, CT, USA. ³⁶Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK. ³⁷Department of Entomology, University of California Riverside, Riverside, CA, USA. ³⁸Department of Bioengineering, Rice University, Houston, TX, USA. ³⁹Department of Pediatrics, Texas Children's Hospital, Houston, TX, USA. ⁴⁰Department of Entomology, Center for Disease Vector Research and Institute for Integrative Genome Biology, University of California, Riverside, CA, USA. ⁴¹Department of Entomology, Virginia Tech, Blacksburg, VA, USA. ⁴²Institute of Integrative Biology, University of Liverpool, Liverpool, UK. ⁴³Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA. ⁴⁴Tata Institute for Genetics and Society, University of California, San Diego, La Jolla, CA, USA. ⁴⁵Department of Entomology, Texas A&M University, College Station, TX, USA. ⁴⁶Laboratory of Ecology, Genetics and Environmental Protection, Tomsk State University, Tomsk, Russia. ⁴⁷Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY, USA. ⁴⁸Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ⁴⁹These authors contributed equally: Benjamin J. Matthews, Olga Dudchenko, Sarah B. Kingan. *e-mail: bnmthws@gmail.com

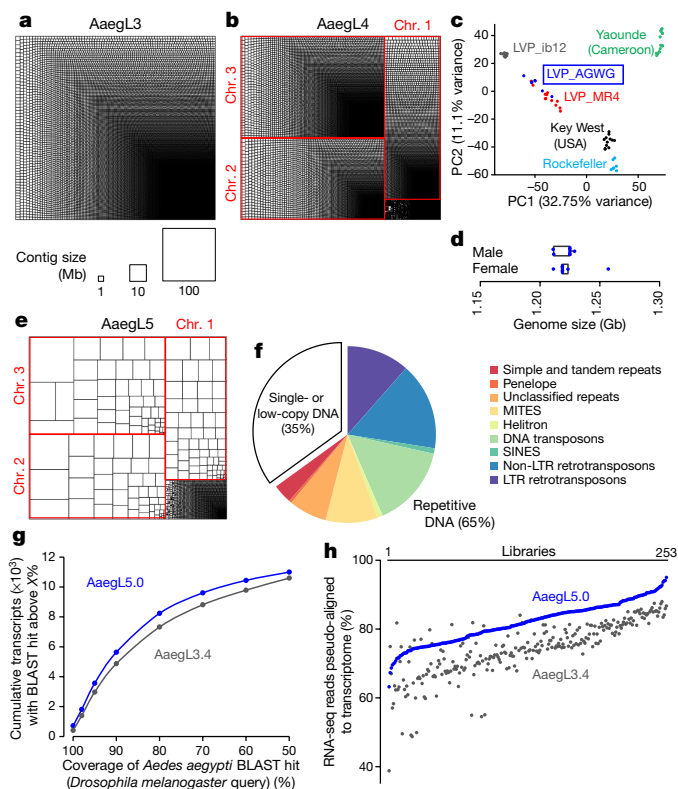


Fig. 1 | AaegL5 assembly statistics and annotation. **a**, **b**, Treemap of AaegL3 (**a**) and AaegL4 (**b**) contigs scaled by length. **c**, Principal component analysis of allelic variation of the indicated strains at 11,229 SNP loci. $n = 7$ per genotype. **d**, Flow cytometry analysis of LVP_AGWG genome size. $n = 5$ per sex. Box plot: median is indicated by the blue line; boxes show first to third quartiles, whiskers are the $1.5\times$ interquartile interval (Extended Data Fig. 1b). **e**, Treemap of AaegL5 contigs scaled by length. **f**, Genome composition (Supplementary Data 2, 3). **g**, Gene set alignment BLASTp coverage is compared between AaegL3.4 and AaegL5.0, with *D. melanogaster* protein queries. **h**, Alignment of 253 RNA-seq libraries to AaegL3.4 and AaegL5.0 gene set annotations (Supplementary Data 4–9). LTR, long terminal repeat retrotransposon; MITEs, miniature inverted-repeat transposable elements; SINES, short interspersed nuclear elements.

93% in the number of contigs, and anchored end-to-end to the three *Ae. aegypti* chromosomes (Fig. 1 and Extended Data Figs. 1, 2). Using optical mapping and linked-read sequencing, we validated the local structure and predicted structural variants between haplotypes. We generated an improved gene set annotation (AaegL5.0), as assessed by a mean increase in RNA-sequencing (RNA-seq) read alignment

of 12%, connections between many gene models that were previously split across multiple contigs, and a roughly twofold increase in the enrichment of assay for transposase-accessible chromatin using sequencing (ATAC-seq) alignments near predicted transcription start sites. We demonstrate the utility of AaegL5 and the AaegL5.0 annotation by investigating a number of scientific questions that could not be addressed with the previous genome annotations.

This project used the Liverpool *Aedes* Genome Working Group (LVP_AGWG) strain, related to the AaegL3 Liverpool ib12 (LVP_ib12) assembly strain² (Fig. 1c and Extended Data Fig. 1a). Using flow cytometry, we estimated that the genome size of LVP_AGWG is approximately 1.22 Gb (Fig. 1d and Extended Data Fig. 1b). To generate our primary assembly, we produced 166 Gb of Pacific Biosciences data (around $130\times$ coverage for a 1.28-Gb genome) and assembled the genome using FALCON-Unzip⁵. This resulted in a total assembly length of 2.05 Gb (contig N50, 0.96 Mb; and NG50, 1.92 Mb, meaning half of the expected genome size found on contigs >1.92 Mb). FALCON-Unzip annotated the resulting contigs as either primary (3,967 contigs; N50, 1.30 Mb; NG50, 1.91 Mb) or haplotigs (3,823 contigs; N50, 193 kb), representing alternative haplotypes present in the approximately 80 male siblings pooled for sequencing (Table 1 and Extended Data Fig. 1e). The primary assembly was longer than expected for a haploid *Ae. aegypti* genome, as predicted by flow cytometry and prior assemblies, which was consistent with remaining alternative haplotypes that were too divergent to be automatically identified as primary and associated alternative haplotig pairs.

To generate a linear chromosome-scale reference genome assembly, we combined the primary contigs and haplotigs that were generated by FALCON-Unzip to create an assembly comprising 7,790 contigs. We used Hi-C to order and orient these contigs, correct misjoined sections and merge overlaps (Extended Data Fig. 1c–e). We set aside 359 contigs that were shorter than 20 kb and used the Hi-C data to identify 258 misjoined sections, resulting in 8,306 ordered and oriented contigs. This procedure revealed extensive sequence overlap among the contigs, consistent with the assembly of numerous alternative haplotypes. We developed a procedure to merge these alternative haplotypes, removing 5,440 gaps and boosting the contiguity (N50, 5.0 Mb; NG50, 4.6 Mb). This procedure placed 94% of sequenced (non-duplicated) bases onto three chromosome-length scaffolds that correspond to the three *Ae. aegypti* chromosomes. After scaffolding, we performed gap-filling and polishing using Pacific Biosciences reads. This removed 270 gaps and further increased the contiguity (N50, 11.8 Mb; NG50, 11.8 Mb), resulting in a final 1.279-Gb AaegL5 assembly and a complete mitochondrial genome (Fig. 1e and Table 1). We used Hi-C contact maps to estimate the position of the centromere with a resolution of around 5 Mb: chromosome 1, approximately 150–154 Mb; chromosome 2, around 227–232 Mb, chromosome 3, around 196–201 Mb. There are 229 remaining gaps in the primary assembly, including 173 on the three primary chromosomal scaffolds (Extended Data Fig. 2a and

Table 1 | Comparison of assembly statistics

	Genome assembly			
	AaegL3	AaegL4	AaegL5 FALCON-Unzip	AaegL5 (NCBI) FALCON-Unzip + Hi-C + polish
Total length (non-N bp)	1,310,092,987	1,254,548,160	1,695,064,654	1,278,709,169
Contig number	36,205	37,224	3,967	2,539
Contig N50 (bp)	82,618	84,074	1,304,397	11,758,062
Contig NG50 (bp)	85,043	81,911	1,907,936	11,758,062
Scaffold number	4,757	6,206	N/A	2,310
Scaffold N50 (bp)	1,547,048	404,248,146 ^a	N/A	409,777,670 ^a
GC content (%)	38.27	38.28	38.16	38.18
Alternative haplotypes (bp)	N/A	N/A	351,566,101	591,941,260
Alternative haplotypes (contigs)	N/A	N/A	3,823	4,224

N/A, not applicable.

^aScaffold N50 is the length of chromosome 3.

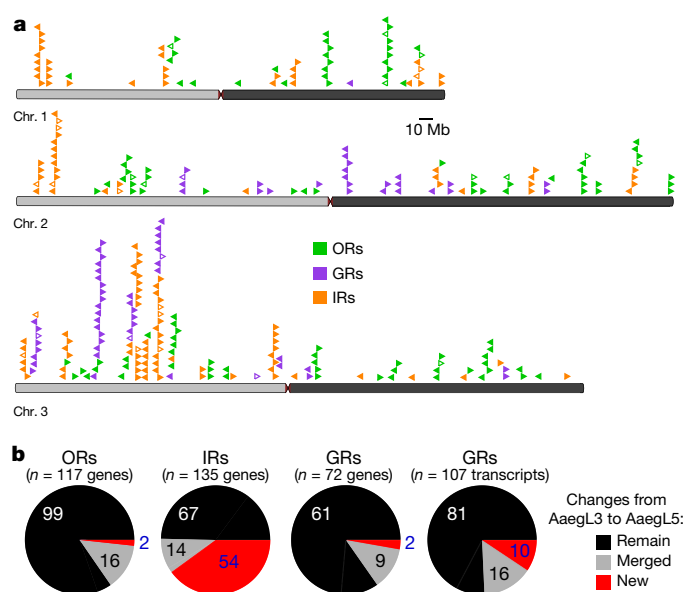


Fig. 2 | Chromosomal arrangement and increased number of chemosensory receptor genes. **a**, Location of predicted chemoreceptors (odorant receptors (ORs), gustatory receptors (GRs) and ionotropic receptors (IRs)) by chromosome in *Ae. aegypti*. The blunt end of the arrowheads marks gene position and the arrow indicates orientation. Filled and open arrowheads represent intact genes and pseudogenes, respectively (Supplementary Data 17–20 and Extended Data Fig. 3). **b**, Chemosensory receptor annotations are compared between *Ae. aegypti* and *Ae. albopictus*.

Supplementary Data 1). Analysis of near-universal single-copy orthologues using BUSCO⁶ revealed a slight increase in complete single-copy orthologues and a reduction in fragmented and missing genes compared to previous assemblies (see Supplementary Methods and Supplementary Discussion). *Ae. aegypti* is markedly more contiguous than *Ae. albopictus* and *Ae. tritaeniorhynchus* assemblies^{2,4} (Fig. 1a, b, e and Table 1). Using the TEfam, Repbase and de novo identified repeat databases, we found that 65% of *Ae. aegypti* was composed of transposable elements and other repetitive sequences (Fig. 1f and Supplementary Data 2, 3).

Complete and correct gene models are essential for studying all aspects of mosquito biology. We used the NCBI RefSeq annotation pipeline to produce annotation version 101 (*Ae. aegypti*; Extended Data Fig. 2b) followed by manual curation of key gene families. *Ae. aegypti* formed the basis for a comprehensive quantification of transcript abundance in 253 sex-, tissue- and developmental stage-specific RNA-seq libraries (Supplementary Data 4–8). The *Ae. aegypti* gene set is considerably more complete and correct than previous versions. Many more genes have high protein coverage when compared to *Drosophila melanogaster* orthologues (915 more genes with >80% coverage, a 12.5% increase over *Ae. albopictus*; Fig. 1g) and >12% more RNA-seq reads map to the *Ae. aegypti* gene set annotation than *Ae. albopictus* (Fig. 1h and Supplementary Data 9). In addition, 1,463 genes that were previously annotated separately as paralogues were collapsed into single gene models and 481 previously fragmented gene models were completed (Supplementary Data 10, 11). For example, *sex peptide receptor* is represented by a six-exon gene model in *Ae. aegypti* compared to two partial gene fragments on separate scaffolds in *Ae. albopictus* (Extended Data Fig. 2c). Genome-wide, we mapped a 1.8-fold higher number of ATAC-seq reads, known to co-localize with promoters and other *cis*-regulatory elements⁷, to predicted transcription start sites in *Ae. aegypti* compared to *Ae. albopictus*, consistent with more complete gene models in *Ae. aegypti* (Extended Data Fig. 2d).

We next validated the base-level and structural accuracy of the *Ae. aegypti* assembly. We estimate the lower bound of base-level accuracy of the assembly to have a quality value of 34.75 (meaning that 99.9665% of bases are correct, see Supplementary Methods and Supplementary

Discussion). To develop a fine-scale physical genome map based on *Ae. aegypti*, we compared the assembly coordinates of 500 bacterial artificial chromosome (BAC) clones containing *Ae. aegypti* genomic DNA with physical mapping by fluorescence in situ hybridization (FISH) (Extended Data Fig. 2e and Supplementary Data 12). After removing repetitive BAC-end sequences and those with ambiguous FISH signals, 377 out of 387 (97.4%) of probes showed concordance between physical mapping and BAC-end alignment. The 10 remaining discordant signals were not supported by Bionano or 10X analysis, and therefore probably do not reflect misassemblies in *Ae. aegypti*. The genome coverage of this physical map is 93.5%, compared to 45% of *Ae. albopictus*^{9,10}, and is one of the most complete genome maps across mosquito species^{9,10}.

Curation of multi-gene families

Large multi-gene families are very difficult to assemble and correctly annotate, because recently duplicated genes typically share high sequence similarities or can be misclassified as alleles of a single gene. We curated genes in large multi-gene families that encode proteases, G protein-coupled receptors, and chemosensory receptors using the improved *Ae. aegypti* genome and *Ae. aegypti* annotation. Serine proteases mediate immune responses¹¹ and metalloproteases have been linked to vector competence and mosquito–*Plasmodium* interactions¹². Gene models for over 50% of the 404 annotated serine proteases and metalloproteases in *Ae. albopictus* were improved in *Ae. aegypti*, and we found 49 previously unannotated protease genes (Supplementary Data 13). G protein-coupled receptors are membrane proteins that respond to diverse external and internal sensory stimuli. We provide major corrections to gene models that encode 10 visual opsins and 17 dopamine and serotonin receptors (Extended Data Fig. 2f and Supplementary Data 14–16). Three large multi-gene families of insect chemosensory receptors are ligand-gated ion channels: odorant receptors (OR gene family), gustatory receptors (GR gene family) and ionotropic receptors (IR gene family). These collectively allow insects to sense a vast array of chemical cues, including carbon dioxide and human body odours that activate and attract female mosquitoes. We identified 117 odorant receptors, 72 gustatory receptors (encoding 107 transcripts) and 135 ionotropic receptors in the *Ae. aegypti* assembly (Fig. 2a, b, Extended Data Fig. 3 and Supplementary Data 17–20), inferred new phylogenetic trees for each family to investigate the relationship of these receptors in *Ae. aegypti*, *Anopheles gambiae* malaria mosquitoes and *D. melanogaster* (Extended Data Figs. 4–6), and revised expression estimates for adult male and female neural tissues using deep RNA-seq¹³ (Extended Data Fig. 7). Our annotation identified 54 new ionotropic receptor genes (Fig. 2b, Extended Data Fig. 3 and Supplementary Data 17), nearly doubling the known members of this family in *Ae. aegypti*. We additionally reannotated ionotropic receptors in *An. gambiae* and found 64 new genes. In *Ae. aegypti*, chemoreceptors are extensively clustered in tandem arrays (Fig. 2a and Extended Data Fig. 3), in particular on chromosome 3p, in which over a third of all chemoreceptor genes ($n = 111$) are found within a 109-Mb stretch. Although 71 gustatory receptor genes are scattered across chromosomes 2 and 3, only *Ae. aegypti*, a subunit of the carbon-dioxide receptor, is found on chromosome 1. Characterization of the full chemosensory receptor repertoire will enable the development of novel strategies to disrupt mosquito biting behaviour.

Structure of the sex-determining M locus

Sex determination in *Aedes* and *Culex* mosquitoes is governed by a dominant male-determining factor (M factor) at a male-determining locus (M locus) on chromosome 1^{14–16}. This chromosome is homomorphic between the sexes except for the M/m karyotype, meaning that males are M/m and females are m/m. Despite the recent discovery of the M factor *Nix* in *Ae. aegypti*¹⁷, which was entirely missing from the previous *Ae. aegypti* genome assemblies^{2,4}, the full molecular properties of the M locus remain unknown. We aligned *Ae. aegypti* (from M/m males) and *Ae. albopictus* (from m/m females), and identified a region that contained *Nix* in *Ae. aegypti* at which the assemblies diverged and that

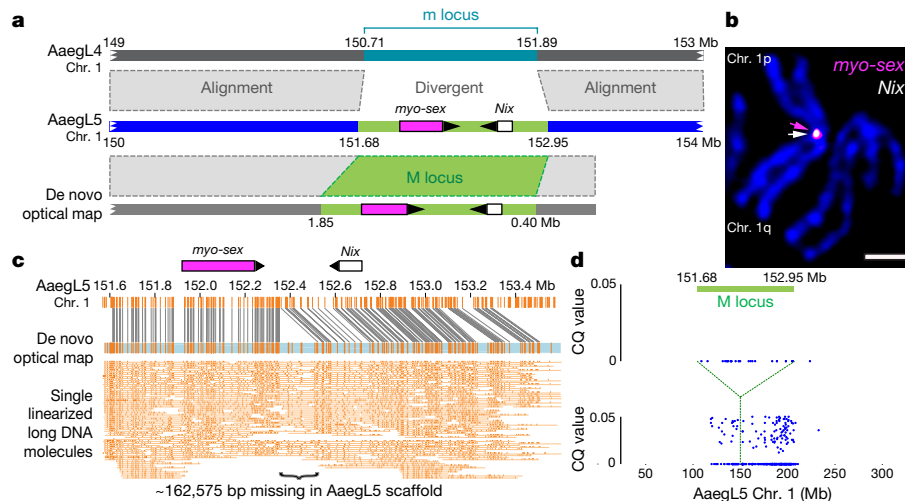


Fig. 3 | Application of AaegL5 to resolve the sex-determining locus. **a**, M locus structure indicating high alignment identity (grey-dashed boxes) and boundaries of *myo-sex* and *Nix* gene models (magenta and white boxes, arrowheads represent orientation). **b**, FISH of BAC clones containing *myo-sex* and *Nix*. Scale bar, 2 μ m. Representative image of 10 samples. **c**, De novo optical map spanning the M locus and bridging the

estimated 163-kb gap in the AaegL5 assembly. DNA molecules are cropped at the edges for clarity. **d**, Chromosome quotient (CQ) analysis of genomic DNA from male and female libraries aligned to AaegL5 chromosome 1. Each dot represents the CQ value of a repeat-masked 1-kb window with >20 reads aligned from male libraries.

may represent the divergent M/m locus (Fig. 3a). A de novo optical map assembly spanned the putative AaegL5 M locus and extended beyond its two borders. We estimated the size of the M locus at approximately 1.5 Mb, including an approximately 163-kb gap between contigs (Fig. 3a, c). We tentatively identified the female m locus as the region in AaegL4 not shared with the M locus-containing chromosome 1, but note that the complete phased structure of the divergent male M locus and corresponding female m locus remain to be determined. *Nix* contains a single intron of 100 kb, while *myo-sex*, a gene encoding a myosin heavy chain protein that has previously been shown to be tightly linked to the M locus¹⁸, is approximately 300 kb in length. More than 73.7% of the M locus is repetitive: long terminal repeat retrotransposons comprise 29.9% of the M locus compared to 11.7% genome-wide. Chromosomal FISH with *Nix*- and *myo-sex*-containing BAC clones¹⁹

showed that these genes co-localize to the 1p pericentromeric region (1p11) in only one homologous copy of chromosome 1, supporting the placement of the M locus at this position in AaegL5 (Fig. 3b). We investigated the differentiation between the sex chromosomes (Fig. 3d) using a chromosome quotient method to quantify regions of the genome with a strictly male-specific signal²⁰. A sex-differentiated region in the LVP_AGWG strain extends to an approximately 100-Mb region surrounding the approximately 1.5-Mb M locus. This is consistent with the recent analysis of male–female F_{ST} in wild population samples and linkage map intercrosses²¹ and could be explained by a large region of reduced recombination encompassing the centromere and M locus²². The availability of a more completely assembled mosquito M locus provides opportunities to study the evolution and maintenance of homomorphic sex-determining chromosomes. The sex-determining

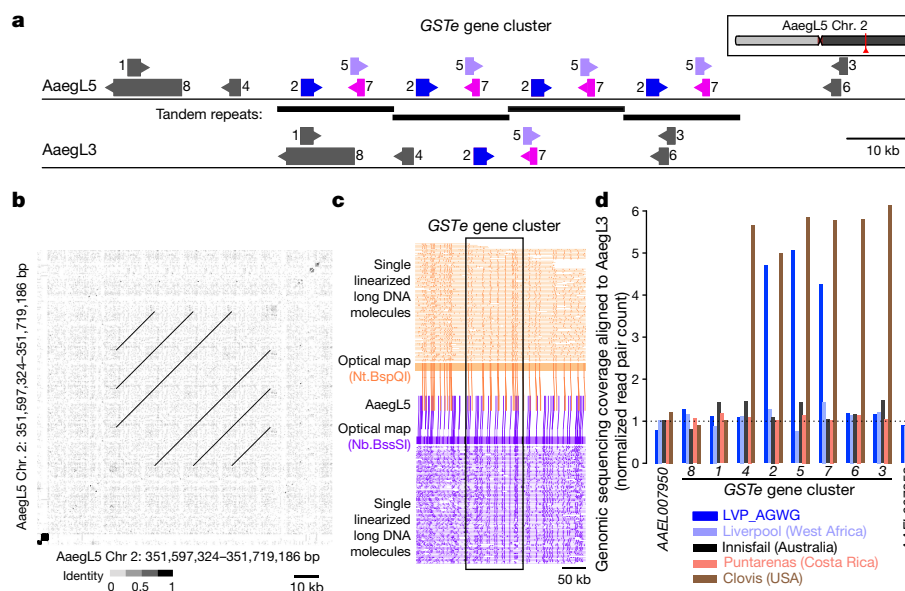


Fig. 4 | Copy-number variation in the glutathione S-transferase epsilon gene cluster. **a**, Glutathione S-transferase epsilon (*GSTe*) gene cluster structure in AaegL5 compared to AaegL3 (Supplementary Data 23). Arrowheads indicate gene orientation. **b**, Dot-plot alignment of AaegL5 *GSTe* region to itself. **c**, Optical mapping of DNA labelled with indicated

enzymes. DNA molecules are cropped at the edges for clarity. **d**, Genomic sequencing coverage of AaegL3 *GSTe* genes (DNA read pairs mapped to each gene, normalized by gene length in kb) from one LVP_AGWG male and pooled mosquitoes from four other laboratory strains.

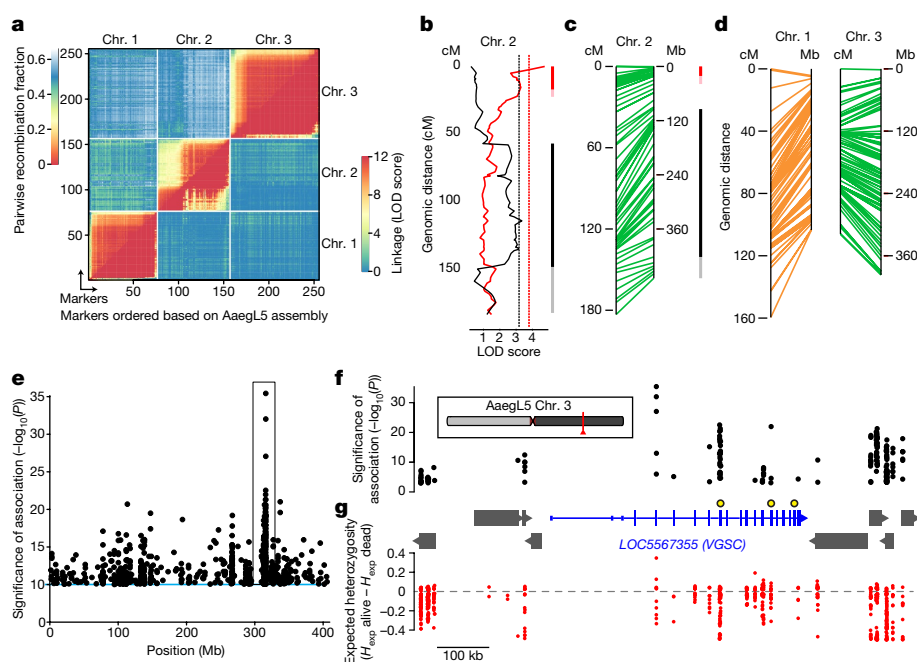


Fig. 5 | Using the Ae. aegypti genome for applied population genetics. **a**, Heat map of linkage based on pairwise recombination fractions for 255 RAD markers ordered by Ae. aegypti physical coordinates. **b**, Significant QTLs on chromosome 2 underlying systemic DENV dissemination in midgut-infected mosquitoes (Extended Data Fig. 10a). Curves represent log of the odds ratio (LOD) scores obtained by interval mapping. Dotted vertical lines indicate genome-wide statistical significance thresholds ($\alpha = 0.05$). Confidence intervals of significant QTLs: bright colour, 1.5-LOD interval; light colour, 2-LOD interval with generalist effects (black, across DENV serotypes and isolates) and DENV isolate-specific effects (red, indicative of genotype-by-genotype interactions). **c**, **d**, Synteny between linkage map (in cM) and physical map (in Mb) for chromosome 2 (**c**) and chromosomes 1

and 3 (**d**). The orange color of chromosome 1 denotes uncertainty in the cM estimates because of deviations in Mendelian ratios surrounding the M locus. **e**, Chromosome 3 SNPs significantly correlated with deltamethrin survival. **f**, **g**, Magnified and inverted view of box in **e**, centred on the new gene model of voltage-gated sodium channel (VGSC, transcript variant X3; the chromosomal position is indicated in red). **f**, Non-coding genes are omitted for clarity, and other genes indicated with grey boxes. VGSC exons are represented by tall boxes and untranslated regions by short boxes. Arrowheads indicate gene orientation. Non-synonymous VGSC SNPs are marked with larger black and yellow circles: V1016I = 315,983,763; F1534C = 315,939,224; V410L = 316,080,722. **g**, Difference in expected heterozygosity (H_{exp} alive – H_{exp} dead) for all SNPs.

chromosome of *Ae. aegypti* may have remained homomorphic at least since the evolutionary divergence between the *Aedes* and *Culex* genera more than 50 million years ago. With the more completely assembled M locus, we can investigate how these chromosomes have avoided the proposed eventual progression into heteromorphic sex chromosomes²³.

Structural variation and gene families

Structural variation is associated with the capacity to vector pathogens²⁴. We produced ‘read cloud’ Illumina sequencing libraries of linked reads with long-range (around 80 kb) phasing information from one male and one female mosquito using the 10X Genomics Chromium platform to investigate structural variants, including insertions, deletions, translocations and inversions, in individual mosquitoes. We observed abundant small-scale insertions and deletions (indels; 26 insertions and 81 deletions called, median 42.9 kb) and inversions and/or translocations (29 called) in these two individuals (Extended Data Fig. 8a and Supplementary Data 21). Eight of the inversions and translocations coincided with structural variants seen independently by Hi-C or FISH, suggesting that those variants are relatively common within this population and can be detected by different methods. Ae. aegypti will provide a foundation for the study of structural variants across *Ae. aegypti* populations.

Hox genes encode highly conserved transcription factors that specify segment identity along the anterior–posterior body axis of all metazoans²⁵. In most vertebrates, *Hox* genes are clustered in a co-linear arrangement, although they are often disorganized or split in other animal lineages²⁶. All expected *Hox* genes are present as a single copy in *Ae. aegypti*, but we identified a split between *labial* and *proboscipedia* placing *labial* on a separate chromosome (Extended Data Fig. 8b and Supplementary Data 22). We confirmed this in Ae. aegypti, which was generated with Hi-C contact maps from a different *Ae. aegypti* strain⁴,

and note a similar arrangement in *Culex quinquefasciatus*, suggesting that it occurred before these two species diverged. Although this is not unprecedented²⁷, a unique feature of this organization is that both *labial* and *proboscipedia* appear to be close to telomeres.

Glutathione S-transferases (GSTs) are a large multi-gene family involved in the detoxification of compounds such as insecticides. Increased GST activity has been associated with resistance to multiple classes of insecticides, including organophosphates, pyrethroids and the organochlorine dichlorodiphenyltrichloroethane (DDT)²⁸. Amplification of detoxification genes is one mechanism by which insects can develop insecticide resistance²⁹. We found that three insect-specific GST epsilon (GSTe) genes on chromosome 2, located centrally in the cluster (*GSTe2*, *GSTe5* and *GSTe7*), are duplicated four times in Ae. aegypti relative to Ae. aegypti (Fig. 4a, b and Supplementary Data 23). Short Illumina read coverage and optical maps confirmed the copy number and arrangement of these duplications in Ae. aegypti (Fig. 4c, d), and analysis of whole-genome sequencing data for four additional laboratory colonies showed variable copy numbers across this gene cluster (Fig. 4d). GSTe2 is a highly efficient metaboliser of DDT³⁰, and it is noteworthy that the cDNA from three GST genes in the quadruplication was detected at higher levels in DDT-resistant *Ae. aegypti* mosquitoes from southeast Asia³¹.

Genome-wide genetic variation

Measurement of genetic variation within and between populations is important for inferring ongoing and historic evolution in a species³². To understand genomic diversity in *Ae. aegypti*, which spread in the last century from Africa to tropical and subtropical regions around the world, we performed whole-genome resequencing on four laboratory colonies. Chromosomal patterns of nucleotide diversity should correlate with regional differences in meiotic recombination rates³³.

We observed pronounced declines in genetic diversity near the centre of each chromosome (Extended Data Fig. 9a, b), providing independent corroboration of the estimated position of each centromere by Hi-C (Extended Data Fig. 2a).

To investigate linkage disequilibrium in geographically diverse populations of *Ae. aegypti*, we first mapped Affymetrix SNP-Chip markers that were designed using AaegL3³⁴ to positions on AaegL5. We genotyped 28 individuals from two populations from Amacuzac, Mexico and Lopé National Park, Gabon and calculated the pairwise linkage disequilibrium of single-nucleotide polymorphisms (SNPs) from 1-kb bins both genome-wide and within each chromosome (Extended Data Fig. 9c, d). The maximum linkage disequilibrium in the Mexican population is approximately twice that of the population from Gabon, which probably reflects a recent bottleneck associated with the spread of this species out of Africa.

Dengue competence and pyrethroid resistance

To illustrate the value of AaegL5 for mapping quantitative trait loci (QTLs), we used restriction site-associated DNA (RAD) markers to locate QTLs underlying dengue virus (DENV) vector competence. We identified and genotyped RAD markers in the F₂ progeny of a laboratory cross between wild *Ae. aegypti* founders from Thailand³⁵ (Extended Data Fig. 10a). For this population, 197 F₂ females had previously been scored for DENV vector competence against four different DENV isolates (two isolates from serotype 1 and two from serotype 3)³⁵. The newly developed linkage map included a total of 255 RAD markers (Fig. 5a) with perfect concordance between genetic distances in centiMorgans (cM) and AaegL5 physical coordinates in Mb (Fig. 5a, c, d). We detected two significant QTLs on chromosome 2 that underlie the likelihood of DENV dissemination from the midgut (that is, systemic infection), an important component of DENV vector competence³⁶. One QTL was associated with a generalist effect across DENV serotypes and isolates, whereas the other was associated with an isolate-specific effect (Fig. 5b, c). QTL mapping powered by AaegL5 will make it possible to understand the genetic basis of *Ae. aegypti* vector competence for arboviruses.

Pyrethroid insecticides are used to combat mosquitoes, including *Ae. aegypti*, and emerging resistance to these compounds is a global problem³⁷. Understanding the mechanisms that underlie insecticide targets and resistance in different mosquito populations is critical to combating arboviral pathogens. Many insecticides act on ion channels, and we curated members of the Cys-loop ligand-gated ion channel (Cys-loop LGIC) superfamily in AaegL5. We found 22 subunit-encoding Cys-loop LGICs (Extended Data Fig. 10d and Supplementary Data 24), of which 14 encode nicotinic acetylcholine receptor (nAChR) subunits. nAChRs consist of a core group of subunit-encoding genes ($\alpha 1$ – $\alpha 8$ and $\beta 1$) that are highly conserved between insect species, and at least one divergent subunit³⁸. Whereas *D. melanogaster* possesses only one divergent nAChR subunit, *Ae. aegypti* has five. We found that agricultural and veterinary insecticides impaired the motility of *Ae. aegypti* larvae (Extended Data Fig. 10c), suggesting that these Cys-loop LGIC-targeting compounds have potential as mosquito larvicides. The improved annotation presented here provides a valuable resource for investigating insecticide efficacy.

To demonstrate how a chromosome-scale genome assembly informs genetic mechanisms of insecticide resistance, we performed a genome-wide population genetic screen for SNPs correlating with resistance to deltamethrin in *Ae. aegypti* collected in Yucatán, Mexico, where pyrethroid-resistant and -susceptible populations co-exist (Fig. 5e). We uncovered an association with non-synonymous changes to three amino acid residues of the voltage-gated sodium channel VGSC, a known target of pyrethroids (Fig. 5f). The gene model for VGSC, a complex locus spanning nearly 500 kb in AaegL5, was incomplete and highly fragmented in AaegL3. SNPs in this region have a lower expected heterozygosity (H_{exp}) in the resistant compared to the susceptible population, suggesting that they are part of a selective sweep for the resistance phenotype surrounding VGSC (Fig. 5g). Accurately

associating SNPs with phenotypes requires a fully assembled genome, and we expect that AaegL5 will be critical to understanding the evolution of insecticide resistance and other important traits.

Summary

The high-quality genome assembly and annotation described here will enable major advances in mosquito biology, and has already allowed us to carry out a number of experiments that were previously impossible. The highly contiguous AaegL5 genome permitted high-resolution genome-wide analysis of genetic variation and the mapping of loci for DENV vector competence and insecticide resistance. A new appreciation of copy number variation in insecticide-detoxifying *GSTe* genes and a more complete accounting of Cys-loop LGICs will catalyse the search for new resistance-breaking insecticides. A doubling in the known number of chemosensory ionotropic receptors provides opportunities to link odorants and tastants on human skin to mosquito attraction, a key first step in the development of novel mosquito repellents. ‘Sterile Insect Technique’ and ‘Incompatible Insect Technique’ show great promise to suppress mosquito populations³⁹, but these population suppression methods require that only males are released. A strategy that connects a gene for male determination to a gene drive construct has been proposed to effectively bias the population towards males over multiple generations⁴⁰, and improved understanding of M locus evolution and the function of its genetic content should facilitate genetic control of mosquitoes that infect many hundreds of millions of people with arboviruses every year¹.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0692-z>.

Received: 28 December 2017; Accepted: 5 October 2018;
Published online 14 November 2018.

- Bhatt, S. et al. The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
- Nene, V. et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **316**, 1718–1723 (2007).
- Timoshevskiy, V. A. et al. An integrated linkage, chromosome, and genome map for the yellow fever mosquito *Aedes aegypti*. *PLoS Negl. Trop. Dis.* **7**, e2052 (2013).
- Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
- Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
- Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2017).
- Denny, S. K. et al. Nf1b promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016).
- Timoshevskiy, V. A. et al. Genomic composition and evolution of *Aedes aegypti* chromosomes revealed by the analysis of physically mapped supercontigs. *BMC Biol.* **12**, 27 (2014).
- George, P., Sharakhova, M. V. & Sharakhov, I. V. High-resolution cytogenetic map for the African malaria vector *Anopheles gambiae*. *Insect Mol. Biol.* **19**, 675–682 (2010).
- Artemov, G. N. et al. The physical genome mapping of *Anopheles albimanus* corrected scaffold misassemblies and identified interarm rearrangements in genus *Anopheles*. *G3 (Bethesda)* **7**, 155–164 (2017).
- Gorman, M. J. & Paskewitz, S. M. Serine proteases as mediators of mosquito immune responses. *Insect Biochem. Mol. Biol.* **31**, 257–262 (2001).
- Goulielmaki, E., Sidén-Kiamos, I. & Loukeris, T. G. Functional characterization of *Anopheles matrix metalloprotease 1* reveals its agonistic role during sporogonic development of malaria parasites. *Infect. Immun.* **82**, 4865–4877 (2014).
- Matthews, B. J., McBride, C. S., DeGennaro, M., Despo, O. & Vossell, L. B. The neurotranscriptome of the *Aedes aegypti* mosquito. *BMC Genomics* **17**, 32 (2016).
- Gilchrist, B. M. & Haldane, J. B. S. Sex linkage and sex determination in a mosquito, *Culex molestus*. *Heredity* **33**, 175–190 (1947).
- McClelland, G. A. H. Sex-linkage in *Aedes aegypti*. *Trans. R. Soc. Trop. Med. Hyg.* **56**, 4 (1962).
- Newton, M. E., Wood, R. J. & Southern, D. I. Cytological mapping of the M and D loci in the mosquito, *Aedes aegypti* (L.). *Genetica* **48**, 137–143 (1978).
- Hall, A. B. et al. A male-determining factor in the mosquito *Aedes aegypti*. *Science* **348**, 1268–1270 (2015).
- Hall, A. B. et al. Insights into the preservation of the homomorphic sex-determining chromosome of *Aedes aegypti* from the discovery of a male-biased gene tightly linked to the M-locus. *Genome Biol. Evol.* **6**, 179–191 (2014).

19. Turner, J. et al. The sequence of a male-specific genome region containing the sex determination switch in *Aedes aegypti*. *Parasit. Vectors* **11**, 549 (2018).
20. Hall, A. B. et al. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **14**, 273 (2013).
21. Fontaine, A. et al. Extensive genetic differentiation between homomorphic sex chromosomes in the mosquito vector, *Aedes aegypti*. *Genome Biol. Evol.* **9**, 2322–2335 (2017).
22. Juneja, P. et al. Assembly of the genome of the disease vector *Aedes aegypti* onto a genetic linkage map allows mapping of genes affecting disease transmission. *PLoS Negl. Trop. Dis.* **8**, e2652 (2014).
23. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118–128 (2005).
24. Riehle, M. M. et al. The *Anopheles gambiae* 2La chromosome inversion is associated with susceptibility to *Plasmodium falciparum* in Africa. *eLife* **6**, e25813 (2017).
25. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
26. Duboule, D. The rise and fall of *Hox* gene clusters. *Development* **134**, 2549–2560 (2007).
27. Negre, B., Ranz, J. M., Casals, F., Cáceres, M. & Ruiz, A. A new split of the *Hox* gene complex in *Drosophila*: relocation and evolution of the gene *labial*. *Mol. Biol. Evol.* **20**, 2042–2054 (2003).
28. Enayati, A. A., Ranson, H. & Hemingway, J. Insect glutathione transferases and insecticide resistance. *Insect Mol. Biol.* **14**, 3–8 (2005).
29. Bass, C. & Field, L. M. Gene amplification and insecticide resistance. *Pest Manag. Sci.* **67**, 886–890 (2011).
30. Ortel, F., Rossiter, L. C., Vontas, J., Ranson, H. & Hemingway, J. Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae*. *Biochem. J.* **373**, 957–963 (2003).
31. Lumjuan, N. et al. The role of the *Aedes aegypti* Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. *Insect Biochem. Mol. Biol.* **41**, 203–209 (2011).
32. Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* **552**, 96–100 (2017).
33. Begun, D. J. & Aquadro, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520 (1992).
34. Evans, B. R. et al. A multipurpose, high-throughput single-nucleotide polymorphism chip for the dengue and yellow fever mosquito, *Aedes aegypti*. *G3 (Bethesda)* **5**, 711–718 (2015).
35. Fansiri, T. et al. Genetic mapping of specific interactions between *Aedes aegypti* mosquitoes and dengue viruses. *PLoS Genet.* **9**, e1003621 (2013).
36. Black, W. C. IV et al. Flavivirus susceptibility in *Aedes aegypti*. *Arch. Med. Res.* **33**, 379–388 (2002).
37. Moyes, C. L. et al. Contemporary status of insecticide resistance in the major *Aedes* vectors of arboviruses infecting humans. *PLoS Negl. Trop. Dis.* **11**, e0005625 (2017).
38. Jones, A. K. & Sattelle, D. B. Diversity of insect nicotinic acetylcholine receptor subunits. *Adv. Exp. Med. Biol.* **683**, 25–43 (2010).
39. Alphey, L. Genetic control of mosquitoes. *Annu. Rev. Entomol.* **59**, 205–224 (2014).
40. Adelman, Z. N. & Tu, Z. Control of mosquito-borne infectious disease: sex and gene drive. *Trends Parasitol.* **32**, 219–229 (2016).

Acknowledgements We thank R. Andino; S. Emrich and D. Lawson (Vectorbase); A. A. James, M. Kunitomi, C. Nusbaum, D. Severson, N. Whiteman; T. Dickinson, M. Hartley and B. Rice (Dovetail Genomics) for early participation in the AGWG; C. Bargmann, D. Botstein, E. Jarvis and E. Lander for encouragement and facilitation. N. Keivanfar, D. Jaffe and D. M. Church (10X Genomics) prepared DNA for structural-variant analysis. We thank A. Harmon of the New York Times and acknowledge generous pro bono data and analysis from our corporate collaborators. This research was supported in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under grant number U19AI110818 to the Broad Institute (S.N.R. and D.E.N.); USDA 2017-05741 (E.L.A.); NSF PHY-1427654 Center for Theoretical Biological Physics (E.L.A.); NIH Intramural Research Program, National Library of Medicine and National Human Genome Research Institute (A.M.P. and S.K.) and the following extramural NIH grants: R01AI101112 (J.R.P.), R35GM118336 (R.S.M. and W.J.G.), R21AI121853 (M.V.S., I.V.S. and A.S.), R01AI123338 (Z.T.), T32GM007739 (M.H.), NIH/NCATS UL1TR000043 (Rockefeller University), DP2OD008540 (E.L.A.), U01AI088647, 1R01AI121211 (W.C.B. IV), Fogarty Training Grant D43TW001130-08, U01HL130010 (E.L.A.), UM1HG009375 (E.L.A.), 5K22AI113060 (O.S.A.), 1R21AI123937 (O.S.A.), and R00DC012069 (C.S.M.); Defence Advanced Research Project Agency: HR0011-17-2-0047 (O.S.A.). Other support was provided by Jane Coffin Childs Memorial Fund (B.J.M.), Center for Theoretical Biological Physics

postdoctoral fellowship (O.D.), Robertson Foundation (L.Z.), and McNair & Welch (Q-1866) Foundations (E.L.A.), French Government's Investissement d'Avenir program, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant ANR-10-LABX-62-IBED to L.L.), Agence Nationale de la Recherche grant ANR-17-ERC2-0016-01 (L.L.), European Union's Horizon 2020 research and innovation program under ZikaPLAN grant agreement no. 734584 (L.L.), Pew and Searle Scholars Programs (C.S.M.), Klingenstein-Simons Fellowship in the Neurosciences (C.S.M.). A.M.W., B.J.W., J.E.C. and S.N.M. were supported by Verily Life Sciences. L.B.V. is an investigator of the Howard Hughes Medical Institute.

Reviewer information Nature thanks S. Celniker, A. G. Clark, R. Waterhouse and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions B.J.M. and L.B.V. conceived the study, coordinated data collection and analysis, designed the figures and wrote the paper with input from all authors. B.J.M. developed and distributed animals and/or DNA of the LVP_AGWG strain. P.P., M.L.S. and J.M. carried out Pacific Biosciences sample preparation and sequencing. S.B.K., R.H., J.K., S.K. and A.M.P. were involved in genome assembly. A.R.H., S.C., J.L. and H.C. carried out Bionano optical mapping. O.D., S.S.B., A.D.O., A.P.A. and E.L.A. carried out Hi-C sample preparation, scaffolding and deduplication. The following authors contributed analysis and data to the indicated figures: B.R.E., A.G.-S. and J.R.P. (Fig. 1c); J.S.J. (Fig. 1d); L.Z. (Fig. 1f); E.C., V.S.J., V.K.K., M.R.M., T.D.M. and B.J.M. (Fig. 1g); I.A., O.S.A., J.E.C., A.M.W., B.J.W., R.G.G.K., S.N.M. and B.J.M. (Fig. 1h); C.S.M., H.M.R., Z.Z., N.H.R. and B.J.M. (Fig. 2); Z.T., M.V.S., I.V.S., A.S., Y.W., J.T., A.C.D., A.R.H. and B.J.M. (Fig. 3); G.D.W., B.J.M., A.R.H., S.B.K., A.M.P. and S.K. (Fig. 4); A.F., I.F., T.F., G.R. and L.L. (Fig. 5a–d); C.L.C., K.S.-R., W.C.B. and B.J.M. (Fig. 5e–g); B.J.M. (Extended Data Fig. 1a); J.S.J. (Extended Data Fig. 1b); O.D., S.S.B., A.D.O., A.P.A. and E.L.A. (Extended Data Fig. 1c, d); S.B.K., J.K., O.D., E.L.A., S.K., A.M.P. and B.J.M. (Extended Data Fig. 1e); A.R.H. and B.J.M. (Extended Data Fig. 2a); E.C., V.S.J., V.K.K., M.R.M., T.D.M. and B.J.M. (Extended Data Fig. 2b); M.H. and B.J.M. (Extended Data Fig. 2c, d); A.S., I.V.S. and M.V.S. (Extended Data Fig. 2e); C.A.B.-S., S.S. and C.A.H. (Extended Data Fig. 2f); C.S.M., H.M.R., Z.Z., N.H.R. and B.J.M. (Extended Data Figs. 3–7); S.N.R. and D.E.N. (Extended Data Fig. 8a); W.J.G., R.S.M., O.D., E.L.A. and B.J.M. (Extended Data Fig. 8b, c); W.J.G. and R.S.M. (Extended Data Fig. 8d); J.E.C., A.M.W., B.J.W., R.G.G.K. and S.N.M. (Extended Data Fig. 9a, b); B.R.E., A.G.-S. and J.R.P. (Extended Data Fig. 9c, d); A.F., I.F., T.F., G.R. and L.L. (Extended Data Fig. 10a, b); G.J.L., A.K.J., V.R., S.D.B., F.A.P. and D.B.S. (Extended Data Fig. 10c, d); A.R.H. (Supplementary Data 1); L.Z. (Supplementary Data 2, 3); I.A., O.S.A., J.E.C., A.M.W., B.J.W., R.G.G.K., S.N.M. and B.J.M. (Supplementary Data 4–9); E.C., V.S.J., V.K.K., M.R.M. and T.D.M. (Supplementary Data 10, 11); A.S., I.V.S. and M.V.S. (Supplementary Data 12); S.R. and A.S.R. (Supplementary Data 13); C.A.B.-S., S.S. and C.A.H. (Supplementary Data 14–16); C.S.M., H.M.R., Z.Z., N.H.R. and B.J.M. (Supplementary Data 17–20); S.N.R. and D.E.N. (Supplementary Data 21); W.J.G. and R.S.M. (Supplementary Data 22); G.D.W. and B.J.M. (Supplementary Data 23); G.J.L., A.K.J., V.R., S.D.B., F.A.P. and D.B.S. (Supplementary Data 24).

Competing interests P.P., M.L.S., J.M., S.B.K., R.H. and J.K. are employees of Pacific Biosciences, a company developing single-molecule sequencing technologies. J.L., S.C., H.C. and A.R.H. are employees of Bionano Genomics and own company stock options. O.D., S.S.B., A.D.O., A.P.A. and E.L.A. are inventors on a US provisional patent application 62/347,605, filed 8 June 2016, by the Baylor College of Medicine and the Broad Institute.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0692-z>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0692-z>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to B.J.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ethics information. The participation of one human subject in blood-feeding mosquitoes was approved and monitored by The Rockefeller University Institutional Review Board (IRB protocol LVO-0652). This subject gave their written and informed consent to participate.

Mosquito rearing and DNA preparation. *Ae. aegypti* eggs from a strain labelled 'LVP_ib12' were supplied by M.V.S. from a colony maintained at Virginia Tech. We performed a single pair cross between a male and female individual to generate material for Hi-C, Bionano optical mapping, flow cytometry, SNP-Chip analysis of strain variance, paired-end Illumina sequencing and 10X Genomics linked reads (Extended Data Fig. 1a). The same single male was crossed to a single female in two additional generations to generate high-molecular weight (HMW) genomic DNA for Pacific Biosciences long-read sequencing and to establish a colony (LVP_AGWG). Rearing was performed as previously described¹³ and all animals were offered a human arm as a blood source.

SNP analysis of mosquito strains. Data were generated as described³⁴, and PCA was performed using LEA 2.0 available for R v.3.4.0^{41,42}. The following strains were used: *Ae. aegypti* LVP_AGWG (samples from the laboratory strain used for the AaegL5 genome assembly, reared as described in Extended Data Fig. 1a by a single pair mating in 2016 from a strain labelled LVP_ib12 maintained at Virginia Tech), *Ae. aegypti* LVP_ib12 (laboratory strain, LVP_ib12, provided in 2013 by D. Severson, University of Notre Dame), *Ae. aegypti* LVP_MR4 (laboratory strain labelled LVP_ib12 obtained in 2016 from MR4 at the Centers for Disease Control via BEI Resources catalogue MRA-735), *Ae. aegypti* Yaounde, Cameroon (field specimens collected in 2014 and provided by B. Kamgang), *Ae. aegypti* Rockefeller (laboratory strain provided in 2016 by G. Dimopoulos, Johns Hopkins Bloomberg School of Public Health), *Ae. aegypti* Key West, Florida (field specimens collected in 2016 and provided by W. Tabachnick). Strains used for the linkage disequilibrium data presented in Extended Data Fig. 9c, d were: *Ae. aegypti* from Amacuzac, Morelos, Mexico (field specimens collected in 2016 and provided by C. Gonzalez Acosta) and *Ae. aegypti* from La Lope National park forest, Gabon (field specimens collected and provided by S. Xia).

Flow cytometry. Genome size was estimated by flow cytometry as described⁴³, except that the propidium iodide was added at a concentration of 25 $\mu\text{L mg}^{-1}$, not 50 $\mu\text{L mg}^{-1}$, and samples were stained in the cold and dark for 24 h to allow the stain to fully saturate the sample. In brief, nuclei were isolated by placing a single frozen head of an adult sample along with a single frozen head of an adult *Drosophila virilis* female standard from a strain with $1\text{C} = 328\text{ Mb}$ into 1 ml of Galbraith buffer (4.26 g MgCl_2 , 8.84 g sodium citrate, 4.2 g 3-[N-morpholino] propane sulfonic acid (MOPS), 1 ml Triton X-100 and 1 mg boiled RNase A in 1 l of ddH₂O, adjusted to pH 7.2 with HCl and filtered through a 0.22- μm filter)⁴⁴ and grinding with 15 strokes of the A pestle at a rate of 3 strokes per 2 s. The resultant ground mixture was filtered through a 60- μm nylon filter (Spectrum Labs). Samples were stained with 25 μg of propidium iodide and held in the cold (4 °C) and dark for 24 h at which time the relative red fluorescence of the 2C nuclei of the standard and sample were determined using a Beckman Coulter CytoFlex flow cytometer with excitation at 488 nm. At least 2,000 nuclei were scored under each 2C peak and all scored peaks had a coefficient of variation of 2.5 or less^{43,44}. Average channel numbers for sample and standard 2C peaks were scored using CytExpert software version 1.2.8.0 supplied with the CytoFlex flow cytometer. Significant differences among strains were determined using Proc GLM in SAS with both a Tukey and a Sheffé option. Significance levels were the same with either option. Genome size was determined as the ratio of the mean channel number of the 2C sample peak divided by the mean channel number of the 2C *D. virilis* standard peak times 328 Mb, where 328 Mb is the amount of DNA in a gamete of the standard. The following species/strains were used: *Ae. mascarensis* (collected by A. Bheecarry on Mauritius in December 2014. Colonized and maintained by J.R.P.), *Ae. aegypti* Ho Chi Minh City F13 (provided by D. J. Gubler, Duke-National University of Singapore as F₁ eggs from females collected in Ho Chi Minh City in Vietnam, between August and September 2013. Colonized and maintained for 13 generations by A.G.-S.), *Ae. aegypti* Rockefeller (laboratory strain provided by D. Severson, Notre Dame), *Ae. aegypti* LVP_AGWG (reared as described in Extended Data Fig. 1a from a strain labelled LVP_ib12 maintained by M.V.S. at Virginia Tech), *Ae. aegypti* New Orleans F8 (collected by D. Wesson in New Orleans 2014, colonized and maintained by J.R.P. through 8 generations of single pair mating), *Ae. aegypti* Uganda 49-ib-G5 (derived by C.S.M. through 5 generations of full-sibling mating of the U49 colony established from eggs collected by J.-P. Mutebi in Entebbe, Uganda in March 2015).

Pacific Biosciences library construction, sequencing and assembly. HMW DNA extraction for Pacific Biosciences sequencing. HMW DNA extraction for Pacific Biosciences sequencing was performed using the Qiagen MagAttract Kit (67563)

following the manufacturer's protocol with approximately 80 male sibling pupae in batches of 25 mg.

SMRTbell library construction and sequencing. Three libraries were constructed using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). In brief, genomic DNA (gDNA) was mechanically sheared to 60 kb using the Megaruptor system (Diagenode) followed by DNA damage repair and DNA end repair. Universal blunt hairpin adapters were then ligated onto the gDNA molecules after which non-SMRTbell molecules were removed with exonuclease. Pulse-field gels were run to assess the quality of the SMRTbell libraries. Two libraries were size-selected using SageELF (Sage Science) at 30 kb and 20 kb, the third library was size-selected at 20 kb using BluePippin (Sage Science). Prior to sequencing, another DNA-damage repair step was performed and quality was assessed with pulse-field gel electrophoresis. A total of 177 SMRT cells were run on the RS II using P6-C4 chemistry and 6 h videos.

Contig assembly and polishing. A diploid contig assembly was carried out using FALCON v.0.4.0 followed by the FALCON-Unzip module (revision 74eefabdc-c4849a8cef24d1a1bbb27d953247bd7)⁵. The resulting assembly contains primary contigs, a partially phased haploid representation of the genome and haplotigs, which represent phased alternative alleles for a subset of the genome. Two rounds of contig polishing were performed. For the first round, as part of the FALCON-Unzip pipeline, primary contigs and secondary haplotigs were polished using haplotype-phased reads and the Quiver consensus caller⁴⁵. For the second round of polishing we used the 'resequencing' pipeline in SMRT Link v.3.1, with primary contigs and haplotigs concatenated into a single reference. Resequencing maps all raw reads to the combined assembly reference with BLASR (v.3.1.0)⁴⁶, followed by consensus calling with Arrow (<https://github.com/PacificBiosciences/GenomicConsensus>)⁴⁶.

Hi-C sample preparation and analysis. *Library preparation.* In brief, insect tissue was crosslinked and homogenized. The nuclei were then extracted and permeabilized, and libraries were prepared using a modified version of the in situ Hi-C protocol that we optimized for insect tissue⁴⁷. Separate libraries were prepared for samples derived from three individual male pupae. The resulting libraries were sequenced to yield 118 million, 249 million and 114 million reads (coverage: 120 \times) and these were processed using Juicer⁴⁸.

Hi-C approach. Using the results of FALCON-Unzip as input, we used Hi-C to correct misjoins, to order and orient contigs, and to merge overlaps (Extended Data Fig. 1c–e). The Hi-C based assembly procedure that we used is described in detail in the Supplementary Methods and Supplementary Discussion. Notably, both primary contigs and haplotigs were used as input. This was essential, because Hi-C data identified genomic loci in which the corresponding sequence was absent in the primary FALCON-Unzip contigs, and present only in the haplotigs; the loci would have led to gaps, instead of contiguous sequence, if the haplotigs were excluded from the Hi-C assembly process (Extended Data Fig. 1e).

Hi-C scaffolding. We set aside 359 FALCON-Unzip contigs shorter than 20 kb, because such contigs are more difficult to accurately assemble using Hi-C. To generate chromosome-length scaffolds, we used the Hi-C maps and the remaining contigs as inputs to the previously described algorithms⁴. Note that both primary contigs and haplotigs were used as input. We performed quality control, manual polishing and validation of the scaffolding results using Assembly Tools⁴⁹. This produced three chromosome-length scaffolds. Notably, the contig N50 decreased slightly, to 929,392 bp, because of the splitting of misjoined contigs.

Hi-C alternative haplotype merging. Examination of the initial chromosome-length scaffolds using Assembly Tools⁴⁹ revealed that extensive undercollapsed heterozygosity was present. In fact, most genomic intervals were repeated, with variations, on two or more unmerged contigs. This suggested that the levels of undercollapsed heterozygosity were unusually high, and that the true genome length was far shorter than either the total length of the Pacific Biosciences contigs (2,047 Mb), or the initial chromosome-length scaffolds (1,973 Mb). Possible factors that could have contributed to the unusually high rate of undercollapsed heterozygosity seen in the FALCON-Unzip Pacific Biosciences contigs relative to prior contig sets for *Ae. aegypti* generated using Sanger sequencing (AaegL3)², include high heterozygosity levels in the species and incomplete inbreeding in the samples that we sequenced. The merge algorithm described previously⁴ detects and merges draft contigs that overlap one another owing to undercollapsed heterozygosity. Because undercollapsed heterozygosity does not affect most loci in a typical draft assembly, the default parameters are relatively stringent. We adopted more permissive parameters for AaegL5 to accommodate the exceptionally high levels of undercollapsed heterozygosity, but found that the results would occasionally merge contigs that did not overlap. To avoid these false positives, we developed a procedure to manually identify and 'whitelist' regions of the genome containing no overlap, based on both Hi-C maps and LASTZ alignments (Extended Data Fig. 1c, Supplementary Methods and Supplementary Discussion). We then reran the merge step, using the whitelist as an additional input. Finally, we performed quality control of the results using Assembly Tools⁴⁹, which confirmed the absence of the undercollapsed

heterozygosity that we had previously observed. The resulting assembly contained three chromosome-length scaffolds (310 Mb, 473 Mb and 409 Mb), which spanned 94% of the merged sequence length. The assembly also contained 2,364 small scaffolds, which spanned the remaining 6% (Table 1). These lengths were consistent with the results of flow cytometry and the lengths obtained in prior assemblies. Notably, the merging of overlapping contigs using the above procedure frequently eliminated gaps, and thus greatly increased the contig N50, from 929,392 to 4,997,917 bp.

Final gap-filling and polishing. *Scaffolded assembly polishing.* Following scaffolding and de-duplication, we performed a final round of arrow polishing. PBJelly⁵⁰ from PBSuite version 15.8.24 was used for gapfilling of the de-duplicated HiC assembly (see 'Protocol.xml' in Supplementary Methods and Supplementary Discussion). After PBJelly, the leftover file was used to translate the renamed scaffolds to their original identifiers. For this final polishing step (run with SMRT Link v3.1.1 resequencing), the reference sequence included the scaffolded, gap-filled reference, as well as all contigs and contig fragments not included in the final scaffolds (https://github.com/skingan/AaegL5_FinalPolish). This reduces the likelihood that reads map to the wrong haplotype, by providing both haplotypes as targets for read mapping. For submission to NCBI, two scaffolds identified as mitochondrial in origin were removed (see below), and all remaining gaps on scaffolds were standardized to a length of 100 Ns to indicate a gap of unknown size. The assembly quality value was estimated using independent Illumina sequencing data from a single individual male pupa (library H2NJHADXY_1/2). Reads were aligned with BWA-MEM v.0.7.12-r1039⁵¹. FreeBayes v.1.1.0-50-g61527c5-dirty⁵² was used to call SNPs and short indels with the parameters -C 2 -O -q 20 -z 0.10 -E 0 -X -u -p 2 -F 0.6. Any SNP and short indels showing heterozygosity (for example, 0/1 genotype) were excluded. The quality value was estimated at 34.75 using the PHRED formula with SNPs as the numerator (597,798) and number of bases with at least threefold coverage as the denominator, including alternate alleles (1,782,885,792). *Identification of mitochondrial contigs.* During the submission process for this genome, two contigs were identified as mitochondrial in origin and were removed from the genomic assembly, manually circularized, and submitted separately. The mitochondrial genome is available as GenBank accession number MF194022.1, RefSeq accession number NC_035159.1.

Bionano optical mapping. *HMW DNA extraction.* HMW DNA extraction was performed using the Bionano Animal Tissue DNA Isolation Kit (RE-013-10), with a few protocol modifications. A single-cell suspension was made as follows. First, 47 mg of frozen male pupae was fixed in 2% v/v formaldehyde in Homogenization Buffer from the kit (Bionano 20278), for 2 min on ice. Then, the pupae were roughly homogenized by blending for 2 s, using a rotor–stator tissue homogenizer (TissueRuptor, Qiagen 9001271). After another 2 min fixation, the tissue was finely homogenized by running the rotor–stator for 10 s. Subsequently, the homogenate was filtered with a 100- μ m nylon filter, fixed with ethanol for 30 min on ice, spun down, and washed with more Homogenization Buffer (to remove residual formaldehyde). The final pellet was resuspended in Homogenization Buffer. A single agarose plug was made using the resuspended cells, using the CHEF Mammalian Genomic DNA Plug Kit (BioRad 170-3591), following the manufacturer's instructions. The plug was incubated with Lysis Buffer (Bionano 20270) and Puregene Proteinase K (Qiagen 1588920) overnight at 50 °C, then again the following morning for 2 h (using new buffer and Proteinase K). The plug was washed, melted and solubilized with GELase (Epicentre G09200). The purified DNA was subjected to 4 h of drop dialysis (Millipore, VCPW04700) and quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen/Molecular Probes P11496).

DNA labelling. DNA was labelled according to commercial protocols using the DNA Labelling Kit NLRs (RE-012-10, Bionano Genomics). Specifically, 300 ng of purified genomic DNA was nicked with 7 U nicking endonuclease Nt.BspQI (New England BioLabs, NEB) at 37 °C for 2 h in NEBuffer3. The nicked DNA was labelled with a fluorescent-dUTP nucleotide analogue using Taq polymerase (NEB) for 1 h at 72 °C. After labelling, the nicks were ligated with Taq ligase (NEB) in the presence of dNTPs. The backbone of fluorescently labelled DNA was counterstained with YOYO-1 (Invitrogen).

Data collection. The DNA was loaded onto the nanochannel array of Bionano Genomics IrysChip by electrophoresis of DNA. Linearized DNA molecules were then imaged automatically followed by repeated cycles of DNA loading using the Bionano Genomics Irys system. The DNA-molecule backbones (YOYO-1 stained) and locations of fluorescent labels along each molecule were detected using the in-house-generated software package, IrysView. The set of label locations of each DNA molecule defines an individual single-molecule map. After filtering data using normal parameters (molecule reads with length greater than 150 kb, a minimum of 8 labels and standard filters for label and backbone signals), a total of 299 Gb and 259 Gb of data were collected from Nt.BspQI and Nb.BssSI samples, respectively.

De novo genome map assembly. De novo assembly was performed with non-haplotype aware settings (optArguments_nonhaplotype_noES_iry.xml) and

pre-release version of Bionano Solve3.1 (Pipeline version 6703 and RefAligner version 6851). On the basis of the overlap–layout–Consensus paradigm, pairwise comparisons of all DNA molecules were performed to create an overlap graph, which was then used to create the initial consensus genome maps. By realigning molecules to the genome maps (RefineB $P = 10 \times 10^{-11}$) and by using only the best match molecules, a refinement step was performed to refine the label positions on the genome maps and to remove chimeric joins. Next, during an extension step, the software aligned molecules to genome maps (extension, $P = 10 \times 10^{-11}$), and extended the maps based on the molecules aligning past the map ends. Overlapping genome maps were then merged using a merge P -value cut-off of $10 P = 10 \times 10^{-15}$. These extension and merge steps were repeated five times before a final refinement was applied to 'finish' all genome maps (refine final, $P = 10 \times 10^{-11}$). Two genome map de novo assemblies, one with nickase Nt.BspQI and the other with nickase Nb.BssSI, were constructed. Alignments between the constructed de novo genome assemblies and the L5 assembly were performed using a dynamic programming approach with a scoring function and a P -value cutoff of $P = 10 \times 10^{-12}$.

Transposable element identification. *Identification of known transposon elements.* We first identified known transposable elements using RepeatMasker (version 3.2.6)⁵³ against the mosquito TEfam (<https://tefam.biochem.vt.edu/tefam/>, data downloaded July 2017), a manually curated mosquito transposable-elements database. We then ran RepeatMasker using the TEfam database and Repbase transposable-elements library (version 10.05). RepeatMasker was set to default parameters with the -s (slow search) flag and NCBI/RMblast program (v.2.2.28). *De novo repeat family identification.* We searched for repeat families and consensus sequences using the de novo repeat prediction tool RepeatModeler (version 1.0.8)⁵⁴ using default parameters with RECON (version 1.07) and RepeatScout (1.0.5) as core programs. Consensus sequences were generated and classified for each repeat family. Then RepeatMasker was run on the genome sequences, using the RepeatModeler consensus sequence as the library.

Tandem repeats. We also predicted tandem repeats in the whole genome and in the repeatmasked genome using Tandem Repeat Finder⁵⁵. Long tandem copies were identified using the 'Match=2, Mismatch=7, Delta=7, PM=80, PI=10, Minscore=50 MaxPeriod=500' parameters. Simple repeats, satellites and low complexity repeats were found using 'Match=2, Mismatch=7, Delta=7, PM=80, PI=10, Minscore=50, and MaxPeriod=12' parameters.

A file representing the coordinates of all identified repeat and transposable-element structures in AaegL5 can be found at <https://github.com/VosshallLab/AGWG-AaegL5>.

Generation of RefSeq gene set annotation. The AaegL5 assembly was deposited at NCBI in June 2017 and annotated using the NCBI RefSeq Eukaryotic gene annotation pipeline⁵⁶. Evidence to support the gene predictions came from over 9 billion Illumina RNA-seq reads, 67,000 Pacific Biosciences IsoSeq transcripts, 300,000 expressed sequence tags and well-supported proteins from *D. melanogaster* and other insects. Annotation Release 101 was made public in July 2017, and specific gene families were subjected to manual annotation and curation. Detailed descriptions of the manual annotation and curation of multigene families (*Hox* genes, proteases, opsins and biogenic amine receptors, chemosensory receptors and LGICs) can be found in the Supplementary Methods and Supplementary Discussion. See also https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Aedes_aegypti/101/.

Alignment of RNA-seq data to AaegL5 and quantification of gene expression. Published RNA-seq reads^{13,57} and unpublished RNA-seq reads from tissue-specific libraries produced by Verily Life Sciences were mapped to the RefSeq assembly GCF_002204515.2_AaegL5.0 with STAR aligner (v.2.5.3a)⁵⁸ using the two-pass approach. Reads were first aligned in the absence of gene annotations using the following parameters: --outFilterType BySJout; --alignIntronMax 1000000; --alignMatesGapMax 1000000; --outFilterMismatchNmax 999; --outFilterMismatchNoverReadLmax 0.04; --clip3pNbases 1; --outSAMstrandField intronMotif; --outSAMAttrIHstart 0; --outFilterMultimapNmax 20; --outSAMAttributes NH HI AS NM MD; --outSAMAttrRGline; --outSAMtype BAM SortedByCoordinate. Splice junctions identified during the first pass mapping of individual libraries were combined and supplied to STAR using the --sjdbFileChrStartEnd option for the second pass. Reads mapping to gene models defined by the NCBI annotation pipeline (GCF_002204515.2_AaegL5.0_genomic.gff) were quantified using featureCounts⁵⁹ with default parameters. Count data were transformed to transcripts per million values using a custom Perl script. Details on libraries, alignment statistics and gene expression estimates (expressed in transcripts per million) are provided as Supplementary Data 4–8.

Identification of 'collapsed' and 'merged' gene models from AaegL3.5 to AaegL5.0. VectorBase annotation AaegL3.5 was compared to NCBI *Ae. aegypti* annotation release 101 on AaegL5.0 using custom code developed at NCBI as part of NCBI's eukaryotic genome annotation pipeline. First, assembly–assembly alignments were generated for AaegL3 (GCA_000004015.3) \times AaegL5.0 (GCF_002204515.2) as part of NCBI's Remap coordinate remapping service,

as described at <https://www.ncbi.nlm.nih.gov/genome/tools/remap/docs/alignments>. The alignments are publicly available in NCBI's Genome Data Viewer (<https://www.ncbi.nlm.nih.gov/genome/gdv/>), the Remap interface, and by FTP in either ASN.1 or GFF3 format (ftp://ftp.ncbi.nlm.nih.gov/pub/remap/Aedes_aegypti/2.1/). Alignments are categorized as either 'first pass' (reciprocity = 3) or 'second pass' (reciprocity = 1 or 2). First pass alignments are reciprocal best alignments, and are used to identify regions on the two assemblies that can be considered equivalent. Second pass alignments are cases in which two regions of one assembly have their best alignment to the same region on the other assembly. These are interpreted to represent regions in which two paralogous regions in AaegL3 have been collapsed into a single region in AaegL5, or vice versa.

For comparing the two annotations, both annotations were converted to ASN.1 format and compared using an internal NCBI program that identifies regions of overlap between gene, mRNA and coding sequence (CDS) features projected through the assembly–assembly alignments. The comparison was performed twice, first using only the first pass alignments, and again using only the second pass alignments corresponding to regions in which duplication in the AaegL3 assembly had been collapsed. Gene features were compared, requiring at least some overlapping CDS in both the old and new annotation to avoid noise from overlapping genes and comparisons between coding and non-coding genes. AaegL5.0 genes that matched to two or more VectorBase AaegL3.5 genes were identified. Matches were further classified as collapsed paralogues if one or more of the matches was through the second pass alignments, or as improvements due to increased contiguity or annotation refinement if the matches were through first pass alignments (for example, two AaegL3.5 genes represent the 5' and 3' ends of a single gene on AaegL5.0, such as *sex peptide receptor*). Detailed lists of merged genes are in Supplementary Data 10, 11.

Comparison of alignment to AaegL3.4 and AaegL5.0. The sequences comprising transcripts from the AaegL5.0 gene set annotation were extracted from coordinates provided in GCF_002204515.2_AaegL5.0_genomic.gtf. Sequences corresponding to AaegL3.4 gene set annotations were downloaded from Vectorbase (<https://www.vectorbase.org/download/aedes-aegypti-liverpooltranscriptsaaegL34fagz>). Salmon (v.0.8.2)⁶⁰ indices were generated with default parameters, and all libraries described in Supplementary Data 4 were mapped to both AaegL3.4 and AaegL5 sequences using 'quant' mode with default parameters. Mapping results are presented as Supplementary Data 9 and Fig. 1h.

ATAC-seq. The previously described ATAC-seq protocol⁶¹ was adapted for *Ae. aegypti* brains. Individual brains from LVP_MR4 non-blood-fed females (Extended Data Fig. 2c, d) or females 48 h or 96 h after taking a human blood meal (data not shown) were dissected in 1 × PBS, immediately placed in 100 µl ice-cold ATAC lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPAL CA-630), and homogenized in a 1.5-ml Eppendorf tube using 50 strokes of a Wheaton 1-ml PTFE-tapered tissue grinder. Animals at 96 h after the blood meal were deprived of access to a water oviposition site and were considered gravid at the time of dissection. Lysed brains were centrifuged at 400g for 20 min at 4 °C and the supernatant was discarded. Nuclei were resuspended in 52.5 µl 1 × Tagmentation buffer (provided in the Illumina Nextera DNA Library Prep Kit) and 5 µl were removed to count nuclei on a haemocytometer. In total, 50,000 nuclei were used for each transposition reaction. The concentration of nuclei in Tagmentation buffer was adjusted to 50,000 nuclei in 47.5 µl Tagmentation buffer and 2.5 µl Tn5 enzyme was added (provided in the Illumina Nextera DNA Library Prep Kit). The remainder of the ATAC-seq protocol was performed as described⁶¹. The final library was purified and size-selected using double-sided AMPure XP beads (0.6 ×, 0.7 ×). The library was checked on an Agilent Bioanalyzer 2100 and quantified using the Qubit dsDNA HS Assay Kit. Resulting libraries were sequenced as 75-bp paired-end reads on an Illumina NextSeq500 platform at an average read depth of 30.5 million reads per sample. Raw fastq reads were checked for nucleotide distribution and read quality using FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and mapped to the AaegL5 and AaegL3 versions of the *Ae. aegypti* genome using Bowtie v.2.2.9⁶². Aligned reads were processed using Samtools 1.3.1⁶³ and Picard 2.6.0 (<http://broadinstitute.github.io/picard/index.html>) and only uniquely mapped and non-redundant reads were used for downstream analyses. To compare the annotation and assembly of the *sex peptide receptor* gene in AaegL3 and AaegL5, we used NCBI BLAST⁶⁴ to identify AAEL007405 and AAEL010313 as gene fragments in AaegL3.4 annotation that map to *sex peptide receptor* in the AaegL5.0 genome (BLAST E values for both queries mapping to *sex peptide receptor* were 0.0). Next, we used GMAP⁶⁵ to align AAEL007405 and AAEL010313 fasta sequences to AaegL5. The resulting GFF3 annotation file was used by Gviz⁶⁶ to plot RNA-seq reads and sashimi plots as well as ATAC-seq reads in the region containing *sex peptide receptor*. Transcription start site analysis was performed using HOMER v.4.9⁶⁷. In brief, databases containing 2-kb windows flanking transcription start sites genome-wide were generated using the 'parseGTF.pl' HOMER script from AaegL3.4 and AaegL5.0 GFF3 annotation files. Duplicate transcription start sites and transcription start sites that were within 20 bp from each other were

merged using the 'mergePeaks' HOMER script. Coverage of ATAC-seq fragments in predicted transcription start site regions was calculated with the 'annotatePeaks.pl' script. Fold change in predicted transcription site regions was calculated by dividing the ATAC fragments per base pair per predicted transcription start site in the AaegL5.0 genome by ATAC fragments per base pair per predicted transcription start site in the AaegL3.4 genome at the 0 base pair point in each predicted transcription start site. Coverage histograms were plotted using ggplot v.2.2.1 in RStudio v.1.1.383, R v.3.4.2⁴².

M locus analysis. *Aligning chromosome assemblies and Bionano scaffolds.* The boundaries of the M locus were identified by comparing the current AaegL5 assembly and the AaegL4 assembly⁴ using a program called LAST⁶⁸ (data not shown). To overcome the challenges of repetitive hits, both AaegL5 and AaegL4 assemblies were twice repeat-masked⁴³ against a combined repeat library of TEFam-annotated transposable elements (<https://tefam.biochem.vt.edu/tefam/>)² and a RepeatModeler output⁵⁴ from the *Anopheles* 16 Genomes project⁶⁹. The masked sequences were then compared using BLASTn⁶⁴ and we then set a filter for downstream analysis to include only alignment with ≥98% identity over 1,000 bp. After the identification of the approximate boundaries of the M locus (and m locus), which contains two male-specific genes, *myo-sex*¹⁸ and *Nix*¹⁷, we zoomed in by performing the same analysis on regions of the M locus and m locus plus 2 Mb flanking regions without repeatmasking. In this and subsequent analyses, only alignment with ≥98% identity over 500 bp were included. Consequently, approximate coordinates of the M locus and m locus were obtained on chromosome 1 of the AaegL5 and AaegL4 assemblies, respectively. Super-scaffold_63 in the Bionano optical map assembly was identified by BLASTN⁶⁴ that spans the entire M locus and extends beyond its two borders.

Chromosome quotient analysis. The chromosome quotient (CQ)²⁰ was calculated for each 1,000-bp window across all AaegL5 chromosomes. To calculate the CQ, Illumina reads were generated from two paired sibling female and male sequencing libraries. To generate libraries for CQ analysis, we performed two separate crosses of a single LVP_AGWG male to 10 virgin females. Eggs from this cross were hatched, and virgin male and female adults collected within 12 h of eclosion to verify their non-mated status. We generated genomic DNA from five males and five females from each of these crosses. Sheared genomic DNA was used to generate libraries for Illumina sequencing with the Illumina TruSeq Nano kit and sequencing performed on one lane of 150-bp paired-end sequencing on an Illumina NextSeq 500 in high-output mode.

For a given sequence S_i of a 1,000-bp window, $CQ_{S_i} = F_{S_i}/M_{S_i}$, where F_{S_i} is the number of female Illumina reads aligned to S_i , and M_{S_i} is the number of male Illumina reads aligned to S_i . Normalization was not necessary for these datasets because the mean and median CQs of the autosomes (chromosomes 2 and 3) are all near 1. A CQ value lower than the 0.05 indicates that the sequences within the corresponding 1,000-bp window had at least 20-fold more hits to the male Illumina data than to the female Illumina data. Not every 1,000-bp window produces a CQ value because many were completely masked by RepeatMasker⁵³. To ensure that each CQ value represents a meaningful data point obtained with sufficient alignments, only sequences with more than 20 male hits were included in the calculation. The CQ values were then plotted against the chromosome location of the 1,000-bp window (Fig. 3d). Under these conditions, there is not a single 1,000-bp fragment on chromosomes 2 and 3 that showed CQ = 0.05 or lower.

Chromosome FISH. Slides of mitotic chromosomes were prepared from imaginal discs of fourth instar larvae following published protocols^{3,70,71}. BAC clones were obtained from the University of Liverpool¹⁹ or from a previously described BAC library⁷². BACs were plated on agar plates (Thermo Fisher) and a single bacterial colony was used to grow an overnight bacterial culture in LB broth plates (Thermo Fisher) at 37 °C. DNA from the BACs was extracted using Sigma PhasePrep TM BAC DNA Kit (Sigma-Aldrich, NA-0100). BAC DNA for hybridization was labelled by nick translation with Cy3-, Cy5-dUTP (Enzo Life Sciences) or Fluorescein 12-dUTP (Thermo Fisher). Chromosomes were counterstained with DAPI in Prolong Gold Antifade (Thermo Fisher). Slides were analysed using a Zeiss LSM 880 Laser Scanning Microscope at 1,000× magnification. We note that localization of the M-locus to 1p11 is supported by both FISH and genomic analyses, but is contrary to a previously published placement at 1q21¹⁷.

Identification and analysis of *Ae. aegypti* GST and P450 genes and validation of the repeat structure of the GSTe cluster. Genes were initially extracted from the AaegL5.0 genome annotation (NCBI release 101) by text search and filtered to remove 'off target' matches (for example, 'cytochrome P450 reductase'), then predicted protein sequences of a small number of representative transcripts were used to search the protein set using BLASTp, to identify by sequence similarity sequences not captured by the text search (resulting in two additional P450s, no GSTs). For each gene family, predicted protein sequences were used to search the proteins of the AaegL3.4 gene set using BLASTp. All best matches, and additional matches with amino acid identity >90% were tabulated for each gene family (Supplementary Data 23) to identify both closely related paralogues and alleles

annotated as paralogues in AaegL3.4. On the basis of a BLASTp search against the AaegL3.4 protein set, the two putative P450 genes not annotated as such in AaegL5.0 (encoding proteins XP_001649103.2 and XP_021694388.1) appear to be incorrect gene models in the AaegL5.0 annotation, which should in fact be two adjacent genes (*CYP9J20* and *CYP9J21* for XP_001649103.2; *CYP6P12* and *CYP6BZ1* for XP_021694388.1). Compared to AaegL3.4, which predicts a single copy each of *GSTe2*, *GSTe5* and *GSTe7*, the NCBI annotation of AaegL5.0 predicts three copies each of *GSTe2* and *GSTe5*, and four copies of *GSTe7*, arranged in a repeat structure. BLASTn searches revealed one additional copy each of *GSTe2* and *GSTe5* in the third duplicated unit. Both contain premature termination codons owing to frameshifts, but these could be owing to uncorrected errors in the assembly. Error correction of all duplicated units was not possible owing to the inability to unequivocally align reads to units not 'anchored' to adjacent single-copy sequence.

To validate these tandem duplications, two lanes of Illumina whole-genome sequence data from a single pupa of the LVP_AGWG strain (H2NJHADXY) were aligned to a hard-masked version of the AaegL3 reference genome using Bowtie2 v.2.2.4⁷³, with '--very-fast-local' alignment parameters, an expected fragment size between 0 and 1,500 bp and relative orientation 'forward-reverse' ('-1 0 -X 1500 -fr'). Aligned reads with a mapping quality less than 10 were removed using Samtools⁶³, 'featureCounts', part of the 'Subread' v.1.5.0-p2 package⁷⁴, was used to assign read pairs or reads ('tags') aligned to either DNA strand ('-s 0') and overlapping the coding regions of a gene by at least 100 bp ('-t CDS-minOverlap 100') to genes as an estimate of representation in the genome. Gene-wise tag counts were normalized by calculating the fragments per kilobase of gene length per million mapped reads (FPKM), using the following equation: (tag count/gene length in kb)/(sum of tag counts for all genes in genome/1,000,000).

Median FPKM for all genes in the genome was calculated (48.22), allowing FPKM of *GSTe* genes to be expressed relative to this. To examine strain differences in coverage at this cluster, we repeated this analysis for the four laboratory colonies analysed in Extended Data Fig. 9a, b. Median FPKM values across all genes ranged from 47.68 to 48.46 and gene-wise FPKM values normalized relative to these medians are plotted in Fig. 4d.

To visualize the sequence identity of the repeat structure in the *GSTe* cluster (Fig. 4b), we extracted the region spanning the cluster from AaegL5 chromosome 2 (351,597,324–351,719,186 bp) and performed alignment of Pacific Biosciences reads using minimap2 v2.1.1 (minimap2 -DP -k7 -w1 -B2 -r200 -g100 -m1 L5_gst.fa L5_gst.fa)⁷⁵ and visualized these alignments using D-GENIES v1.2.0⁹² with minimum identity set to 0.15 and 'Strong Precision' enabled. To validate this repeat structure, we aligned two de novo optical maps created by Bionano using linearized DNA labelled with Nt.BspQI or Nb.BssSI. Single molecules from both maps span the entire region and the predicted restriction pattern provides support for the repeat structure as presented in AaegL5 (Fig. 4c).

QTL mapping of DENV vector competence. In theory, a good-quality genome assembly is not necessary for QTL mapping procedures, because it relies on a linkage map that can be generated de novo from empirical recombination fractions. This typically involves three steps: (i) marker selection based on the Mendelian segregation ratios, (ii) marker assignment to linkage groups and (iii) marker ordering within each linkage group. However, if a high-quality reference genome assembly is available, the physical position of each marker can be determined and this prior information greatly facilitates steps (ii) and (iii), as shown below.

To demonstrate the improvement enabled by our new genome, we generated two linkage maps using the same Illumina sequence data that were aligned either to AaegL3 or AaegL5 genome assemblies. Although the initial number of markers was 616 in both cases, the final linkage map was 3.3-fold denser with AaegL5 than with AaegL3, as shown in Extended Data Fig. 10b. The difference in marker density between the two linkage maps is because many markers were filtered out from the AaegL3 data. Because the AaegL3 assembly is highly fragmented (>4,700 scaffolds), the position of each marker within the linkage groups is primarily determined from the recombination fractions. This ordering step is performed by creating a backbone with a subset of informative markers using a two-point algorithm, followed by the positioning of the remaining markers one at a time using a multi-point method. Only markers that are unambiguously positioned are kept in the final linkage map for QTL mapping. We note that AaegL4, which de-duplicated and scaffolded AaegL3 onto chromosomes⁴, would probably yield a similar improvement in mapping resolution.

Another complication arises for the chromosome 1 in *Ae. aegypti*, because recombination is strongly reduced in the region containing the sex-determining M locus. This leads to the severely biased segregation ratios for markers anchored to this linkage group. In our F₂ intercross design, the fully sex-linked markers lacked the F₀ paternal genotype in F₂ females and segregated in the same manner as a backcross design. No linkage analysis method is readily available to deal with a chromosome that behaves like a mixture of intercross and backcross designs. Therefore, AaegL3-guided linkage analysis and QTL mapping for chromosome 1 were restricted to the fully sex-linked region based on a backcross design.

By contrast, AaegL5-guided linkage analysis and QTL mapping for chromosome 1 made use of all markers regardless of their segregation ratios, allowing chromosome-wide coverage. As mentioned in the present manuscript, the only caveat is that our analytical procedure assumes autosomal Mendelian proportions, which may have resulted in over- or underestimation of linkage distances between markers on chromosome 1. The linkage map was iteratively refined by checking for misplaced markers based on visual inspection of the LOD/rf matrix.

Ultimately, AaegL5 has a markedly improved QTL mapping resolution over AaegL3. For instance, we mapped the same QTL underlying systemic DENV dissemination at the extremity of chromosome 2 with both AaegL3 and AaegL5. The 1.5 LOD support interval was much larger for the AaegL3-guided linkage map (0–50 cM, 74% of the linkage group) than for the AaegL5-guided linkage map (0–17 cM, 9% of the linkage group). We present this analysis in Extended Data Fig. 10b.

Mosquito crosses. A large F₂ intercross was created from a single mating pair of field-collected F₀ founders. Wild mosquito eggs were collected in Kamphaeng Phet Province, Thailand in February 2011 as previously described³⁵. In brief, F₀ eggs were allowed to hatch in filtered tap water and the larvae were reared until the pupae emerged in individual vials. *Ae. aegypti* adults were identified by visual inspection and maintained in an insectary under controlled conditions (28 ± 1 °C, 75 ± 5% relative humidity and 12:12-h light:dark cycle) with access to 10% sucrose. The F₀ male and female initiating the cross were chosen from different collection sites to avoid creating a parental pair with siblings from the same wild mother^{76,77}. Their F₁ offspring were allowed to mass-mate and collectively oviposit to produce the F₂ progeny (Extended Data Fig. 10a). A total of 197 females of the F₂ progeny were used as a mapping population to generate a linkage map and detect QTLs underlying vector competence for DENV.

Vector competence. Four low-passaged DENV isolates were used to orally challenge the F₂ females as previously described³⁵. In brief, four random groups of females from the F₂ progeny were experimentally exposed to two virus isolates belonging to DENV serotype 1 (KDH0026A and KDH0030A) and two virus isolates belonging to DENV serotype 3 (KDH0010A and KDH0014A), respectively. All four virus isolates were derived from human serum specimens collected in 2010 from clinically ill patients who were infected with DENV at the Kamphaeng Phet Provincial Hospital³⁵. Because the viruses were isolated in the laboratory cell culture, informed consent of the patients was not necessary for the present study. Complete viral genome sequences were previously deposited into GenBank (accession numbers HG316481, HG316582, HG316483, and HG316484)³⁵. Phylogenetic analysis assigned the viruses to known viral lineages that were circulating in south-east Asia in the previous years³⁵. Each isolate was amplified twice in C6/36 (*Aedes albopictus*) cell lines (maintained at AFRIMS in Bangkok, Thailand; used only to grow virus, not explicitly authenticated or checked for mycoplasma contamination) before vector competence assays. Four- to seven-day-old F₂ females were starved for 24 h and offered an infectious blood meal for 30 min. Viral titres in the blood meals ranged from 2.0 × 10⁴ to 2.5 × 10⁵ plaque-forming units per ml across all isolates. Fully engorged females were incubated under the conditions described above. Vector competence was scored 14 days after the infectious blood meal according to two conventional phenotypes: (i) midgut infection and (ii) viral dissemination from the midgut. These binary phenotypes were scored based on the presence or absence of infectious particles in body and head homogenates, respectively. Infectious viruses were detected by plaque assay performed in LLC-MK2 (rhesus monkey kidney epithelial) cells as previously described^{35,78}.

Genotyping. Mosquito gDNA was extracted using the NucleoSpin 96 Tissue Core Kit (Macherey-Nagel). For the F₀ male, it was necessary to perform whole-genome amplification using the Repli-g Mini kit (Qiagen) to obtain a sufficient amount of DNA. F₀ parents and females of the F₂ progeny were genotyped using a modified version of the original double-digest restriction site-associated DNA (RAD) sequencing protocol⁷⁹, as previously described⁸⁰. The final libraries were spiked with 15% PhiX and sequenced on an Illumina NextSeq 500 platform using a 150-cycle paired-end chemistry (Illumina). A previously developed bash script pipeline⁸⁰ was used to process the raw sequence reads. High-quality reads (Phred scores >25) trimmed to the 140-bp length were aligned to the AaegL5 reference genome (July 2017) using Bowtie v.0.12.7⁶². Parameters for the ungapped alignment included ≤3 mismatches in the seed, suppression of alignments with >1 best reported alignment under a 'try-hard' option. Variant and genotype calling was performed from a catalogue of RAD loci created with the ref_map.pl pipeline in Stacks v.1.19^{81,82}. Downstream analyses only used high-quality genotypes at informative markers that were homozygous for alternative alleles in the F₀ parents (for example, AA in the F₀ male and BB in the F₀ female), had a sequencing depth ≥10× and were present in ≥60% of the mapping population.

Linkage map. A comprehensive linkage map based on recombination fractions among RAD markers in the F₂ generation was constructed using the R package OneMap v.2.0-3⁸³. Every informative autosomal marker is expected to segregate in the F₂ mapping population at a frequency of 25% for homozygous (AA and

BB) genotypes and 50% for heterozygous (AB) genotypes. Autosomal markers that significantly deviated from these Mendelian segregation ratios ($P < 0.05$) were filtered out using a χ^2 test. Owing to the presence of a dominant male-determining locus on chromosome 1, fully sex-linked markers on chromosome 1 are expected to segregate in F_2 females with equal frequencies (50%) of heterozygous (AB) and F_0 maternal (BB) genotypes, because the F_0 paternal (AA) genotype only occurs in F_2 males. As previously reported²¹, strong deviations from the expected Mendelian segregation ratios were observed for a large proportion of markers assigned to chromosome 1 in the female F_2 progeny. Markers on chromosome 1 were included if they had heterozygous (AB) genotype frequencies inside the 40–60% range and F_0 maternal (BB) genotype frequencies inside the 5%–65% range. These arbitrary boundaries for marker selection were largely permissive for partially or fully sex-linked markers on chromosome 1. Owing to a lack of linkage analysis methods that deal with sex-linked markers when only one sex is genotyped, the recombination fractions between all pairs of selected markers were estimated using the rf.2pts function with default parameters for all three chromosomes. The rf.2pts function, which implements the expectation–maximization (EM) algorithm, was used to estimate haplotype frequencies and recombination rates between markers¹¹ under the assumption of autosomal Hardy–Weinberg proportions. Owing to this analytical assumption, the estimates of cM distances could be over- or underestimated for markers on chromosome 1. Markers linked with a LOD score ≥ 11 were assigned to the same linkage group. Linkage groups were assigned to the three distinct *Ae. aegypti* chromosomes based on the physical coordinates of the AaegL5 assembly. Recombination fractions were converted into genetic distances in cM using the Kosambi mapping function⁸⁴. Linkage maps were exported in the R/qtl environment⁸⁵ in which they were corrected for tight double crossing-overs with the calc.errorlod function based on a LOD cut-off threshold of 4. Duplicate markers with identical genotypes were removed with the findDupMarkers function. To remove markers located in highly repetitive sequences, RAD sequences were blasted against the AaegL5 assembly using BLASTn v.2.6.0. Markers with >1 blast hit on chromosomes over their 140-bp length and 100% identity were excluded from linkage analysis. Reported RAD markers were distributed as follows: chromosome 1, $n = 76$; chromosome 2, $n = 80$; chromosome 3, $n = 99$.

QTL mapping. The newly developed linkage map was used to detect and locate QTLs that underlie the DENV vector competence indices described above. Midgut infection was analysed in all F_2 females whereas viral dissemination was analysed only in midgut-infected females. The four different DENV isolates were included as a covariate to detect QTL–isolate interactions. Single QTL detection was performed in the R/qtl environment⁸⁵ using the expectation–maximization algorithm of the scanone function using a binary trait model. Genome-wide statistical significance was determined by empirical permutation tests, with 1,000 genotype–phenotype permutations of the entire dataset.

Comparison between AaegL5 and AaegL3. To assess the improvement obtained in AaegL5 to perform QTL mapping, a linkage map was built by aligning RAD markers to the AaegL3 assembly. The AaegL3-guided linkage map was built by assigning markers to chromosomes and by ordering them within each linkage group only on the basis of their recombination fractions. Markers were initially filtered based on their segregation ratios as described above and assigned to the same linkage group based on a LOD score threshold of ≥ 14 . Linkage groups were assigned to the three *Ae. aegypti* chromosomes using supercontigs that were previously mapped to the chromosomes²². For each linkage group, a backbone was created with a small subset of informative markers ($n = 6$) using the rf.2pts two-point algorithm of the OneMap package. The remaining markers were positioned one at a time using the OneMap order.seq multi-point method, which compares all maps including the new marker at all possible positions keeping the original linkage map unchanged. This procedure produces both a ‘safe’ and a ‘forced’ marker order. The ‘forced’ marker map indicates the most likely position for each marker, whereas the ‘safe’ marker map only displays the unambiguously positioned markers. The AaegL3-guided QTL mapping was performed with the ‘safe’ marker map. Strong bias in Mendelian segregation ratios of markers anchored to chromosome 1 impeded their ordering. Fully sex-linked markers lacked the F_0 paternal (AA) genotype in F_2 females, and segregated analogously to a backcross design in which F_1 AB heterozygotes are backcrossed to F_0 BB homozygotes. No linkage analysis method is readily available to deal with a chromosome that behaves like a mixture of intercross and backcross designs. Therefore, AaegL3-guided linkage analysis and QTL mapping for chromosome 1 were restricted to the fully sex-linked region based on a backcross design. A new OneMap input file only including markers lacking the F_0 paternal (AA) genotype was made by setting the population type to ‘backcross’ instead of ‘F2 intercross’. Markers were ordered using the order.seq function of the OneMap package as described above. A table summarizing this comparison is available as Extended Data Fig. 10b.

Mapping insecticide resistance and VGSC. The mosquito population Viva Cauceal from Yucatán State in Southern Mexico (longitude -89.71827 , latitude 20.99827), was collected in 2011 through the Universidad Autónoma de Yucatán. We identified

up to 25 larval breeding sites from 3–4 city blocks and collected around 1,000 larvae. Larvae were allowed to eclose, and twice a day we aspirated the adults from the cartons, discarding anything other than *Ae. aegypti*. Then, 300–400 *Ae. aegypti* were released into a 2-foot (61-cm) cubic cage where they were allowed to mate for up to five days with ad libitum access to sucrose, after which they were blood fed to collect eggs for the next generation. Then, 390 adult mosquitoes were phenotyped for deltamethrin resistance. We exposed groups of 50 mosquitoes (3–4 days old) to 3 μ g of deltamethrin-coated bottles for 1 h. After this time, active mosquitoes were transferred to cardboard cups and placed into an incubator (28 °C and 70% humidity) for 4 h; these mosquitoes were referred to as the resistant group. Immobility mosquitoes were transferred to a second cardboard cup. After 4 h, newly recovered mosquitoes were aspirated, frozen and labelled as recovered; these were excluded from the current study. The mosquitoes that were immobilized and remained inactive at 4 h post-treatment were scored as susceptible. DNA was isolated from individual mosquitoes by the salt extraction method⁸⁶ and resuspended in 150 μ l of TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0). We constructed a total of four gDNA libraries. Two groups were pooled from DNA of 25 individual females that survived 1 h of deltamethrin exposure (resistant replicates 1 and 2). The second set of two libraries was obtained by pooling DNA from 25 females that were immobilized and inactive at 4 h post-treatment (susceptible replicates 1 and 2). Before pooling, DNA from each individual mosquito was quantified using the Quant-IT Pico Green kit (Life Technologies, Thermo Fisher Scientific) and around 40 ng from each individual DNA sample (25 individuals per library) was used for a final DNA pool of 1 μ g. Pooled DNA was sheared and fragmented by sonication to obtain fragments between 300 and 500 bp (Covaris). We prepared one library for each of the four DNA pools following the Low Sample protocol from the Illumina TrueSeqDNA PCR-Free Sample preparation guide (Illumina). Because 65% of the *Ae. aegypti* genome consists of repetitive DNA, we performed an exome-capture hybridization to enrich for coding sequences using custom SeqCap EZ Developer probes (NimbleGen, Roche). Probes covered protein-coding sequences (not including untranslated regions) in the AaegL1.3 genebuild using previously specified exonic coordinates⁸⁷. In total, 26.7 Mb of the genome (2%) was targeted for enrichment. TruSeq libraries were hybridized to the probes using the xGenLockDown recommendations (Integrated DNA Technologies). The targeted DNA was eluted and amplified (10–15 cycles) before being sequenced on one flow cell of a 100-bp HiSeq Rapid-duo paired-end sequencing run (Illumina) performed by the Centers for Disease Control (Atlanta, GA, USA).

The raw sequence files (FASTQ) for each pair-ended gDNA library were aligned to a custom reference physical map generated from the assembly AaegL5. Nucleotide counts were loaded into a contingency table with four rows corresponding to ‘Alive Rep1’, ‘Alive Rep2’, ‘Dead Rep1’ and ‘Dead Rep2’. The numbers of columns (c) corresponded to the number of alternative nucleotides at a SNP locus. The maximum value for c is 6, corresponding to A, C, G, T, insert or deletion. Three ($2 \times c$) contingency tables were subjected to χ^2 analyses ($c - 1$ degrees of freedom) to determine whether there are significant ($P \leq 0.05$) differences between (1) Alive replicates, (2) Dead replicates and (3) Alive versus Dead. If analysis (1) or (2) was significant, then that SNP locus was discarded. Otherwise the third contingency table consisted of two rows corresponding to Alive (sum of replicates 1 and 2), Dead (replicates 1 and 2 summed), and c columns. The χ^2 value from the ($2 \times c$) contingency χ^2 analysis with ($c - 1$) degrees of freedom was loaded into Excel to calculate the one-tailed probability of the χ^2 distribution probability (P). This value was transformed with $-\log_{10}(P)$. The experiment-wise error rate was then calculated following the method of Benjamini and Hochberg⁸⁸ to lower the number of type I errors (false positives).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

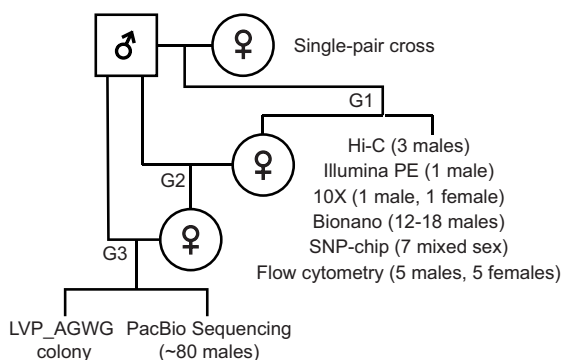
Code availability. The overview of the Hi-C workflow, as well as modifications to 3D-DNA associated with AaegL5, is shared on GitHub at <https://github.com/theaidenlab/AGWG-merge>. The source code and executable version of Juicebox Assembly Tools are available at <http://aidenlab.org/assembly>. Data files and scripts used for the final polishing of scaffolded, gap-filled assembly are available at https://github.com/skingan/AaegL5_FinalPolish.

Data availability. All raw data have been deposited at NCBI under the following BioProject accession numbers: PRJNA318737 (primary Pacific Biosciences data, Hi-C sequencing primary data and processed contact maps, whole-genome sequencing data from a single male (Fig. 4d) and pools of male and females (Fig. 3d), Bionano optical mapping data (Figs. 3c, 4c) and 10X linked-read sequences (Extended Data Fig. 8a and Supplementary Data 21)); PRJNA236239 (RNA-seq reads and de novo transcriptome assembly¹³ (Extended Data Fig. 2c and Supplementary Data 4, 5, 7, 9)); PRJNA209388 (RNA-seq reads for developmental time points⁵⁷ (Fig. 1h and Supplementary Data 4–6, 9)); PRJNA419241 (RNA-seq reads from adult reproductive tissues and developmental time points, Verily Life Sciences (Fig. 1h and Supplementary Data 4, 5, 8, 9)); PRJNA393466 (full-length Pacific Biosciences Iso-Seq transcript sequencing); PRJNA418406 (ATAC-seq data

from adult female brains at three points in the gonotrophic cycle (Extended Data Fig. 2c, d and data not shown)); PRJNA419379 (whole-genome sequencing data from four colonies (Fig. 4d and Extended Data Fig. 9a, b)); PRJNA399617 (restriction-site-associated DNA-sequencing data (Fig. 5a–d)); PRJNA393171 (exome-sequencing data (Fig. 5e–g)). Intermediate results related to the AeGL5 assembly are also available via GitHub (<http://github.com/theaidenlab/AGWG-merge>) and have been uploaded to GEO (GSE113256). The Hi-C maps are available via <http://aidenlab.org/juicebox>. The complete mitochondrial genome is available as Genbank accession MF194022.1, RefSeq accession NC_035159.1. The final genome assembly and annotation are available from the NCBI Assembly Resource under accession GCF_002204515.2.

41. Frichot, E. & François, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
42. R Core Team. *R: A Language and Environment for Statistical Computing* <http://www.R-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2017).
43. Hare, E. E. & Johnston, J. S. Genome size determination using flow cytometry of propidium iodide-stained nuclei. *Methods Mol. Biol.* **772**, 3–12 (2012).
44. Galbraith, D. W. et al. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220**, 1049–1051 (1983).
45. Chin, C. S. et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
46. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**, 238 (2012).
47. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
48. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
49. Dudchenko, O. et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. Preprint at <https://www.biorxiv.org/content/early/2018/01/28/254797> (2018).
50. English, A. C. et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **7**, e47768 (2012).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
52. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at <https://arxiv.org/abs/1207.3907> (2012).
53. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open version 4.0 <http://www.repeatmasker.org> (2013–2015).
54. Smit, A. F. A. & Hubley, R. RepeatModeler Open version 1.0. <http://www.repeatmasker.org> (2008–2015).
55. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
56. Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M. & Kitts, P. in *The NCBI Handbook* 2nd edn <http://www.ncbi.nlm.nih.gov/books/NBK169439/> (NIH, Bethesda, 2013).
57. Akbari, O. S. et al. The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3 (Bethesda)* **3**, 1493–1509 (2013).
58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
60. Patro, R., Duggal, G., Love, M. I., Iziray, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
61. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
63. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
65. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
66. Hahne, F. & Ivanek, R. Visualizing genomic data using Gviz and Bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).
67. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
68. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
69. Neafsey, D. E. et al. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* **347**, 1258522 (2015).
70. Timoshevskiy, V. A., Sharma, A., Sharakhov, I. V. & Sharakhova, M. V. Fluorescent in situ hybridization on mitotic chromosomes of mosquitoes. *J. Vis. Exp.* **67**, e4215 (2012).
71. Sharakhova, M. V. et al. Imaginal discs—a new source of chromosomes for genome mapping of the yellow fever mosquito *Aedes aegypti*. *PLoS Negl. Trop. Dis.* **5**, e1335 (2011).
72. Jiménez, L. V., Kang, B. K., deBruyn, B., Lovin, D. D. & Severson, D. W. Characterization of an *Aedes aegypti* bacterial artificial chromosome (BAC) library and chromosomal assignment of BAC clones for physical mapping quantitative trait loci that influence *Plasmodium* susceptibility. *Insect Mol. Biol.* **13**, 37–44 (2004).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
75. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
76. Apostol, B. L., Black, W. C., IV, Reiter, P. & Miller, B. R. Use of randomly amplified polymorphic DNA amplified by polymerase chain reaction markers to estimate the number of *Aedes aegypti* families at oviposition sites in San Juan, Puerto Rico. *Am. J. Trop. Med. Hyg.* **51**, 89–97 (1994).
77. Rašić, G. et al. The queenslandensis and the type form of the dengue fever mosquito (*Aedes aegypti* L.) are genomically indistinguishable. *PLoS Negl. Trop. Dis.* **10**, e0005096 (2016).
78. Thomas, S. J. et al. Dengue plaque reduction neutralization test (PRNT) in primary and secondary dengue virus infections: how alterations in assay conditions impact performance. *Am. J. Trop. Med. Hyg.* **81**, 825–833 (2009).
79. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135 (2012).
80. Rašić, G., Filipović, I., Weeks, A. R. & Hoffmann, A. A. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics* **15**, 275 (2014).
81. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* **1**, 171–182 (2011).
82. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
83. Margarido, G. R., Souza, A. P. & Garcia, A. A. OneMap: software for genetic mapping in outcrossing species. *Heredity* **144**, 78–79 (2007).
84. Kosambi, D. D. in *The Estimation of Map Distances from Recombination Values* Ch. 15 (ed. Ramaswamy, R.) 125–131 (Springer India, New Delhi, 2016).
85. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
86. Black, W. C. & DuTeau, N. M. in *The Molecular Biology of Insect Disease Vectors*. (eds Crampton, J. M. et al.) 361–373 (Springer, Dordrecht, 1997).
87. Juneja, P. et al. Exome and transcriptome sequencing of *Aedes aegypti* identifies a locus that confers resistance to *Brugia malayi* and alters the immune response. *PLoS Pathog.* **11**, e1004765 (2015).
88. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
89. Robertson, H. M. The insect chemoreceptor superfamily is ancient in animals. *Chem. Senses* **40**, 609–614 (2015).
90. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
91. Merabet, S. & Mann, R. S. To be specific or not: the critical relationship between HOX and TALE proteins. *Trends Genet.* **32**, 334–347 (2016).
92. Cabanettes, F. & Klopp, C. D. GENIES: dot plot large genomes in an interactive, efficient and simple way. *Peer J.* **6**, e4958 (2018).

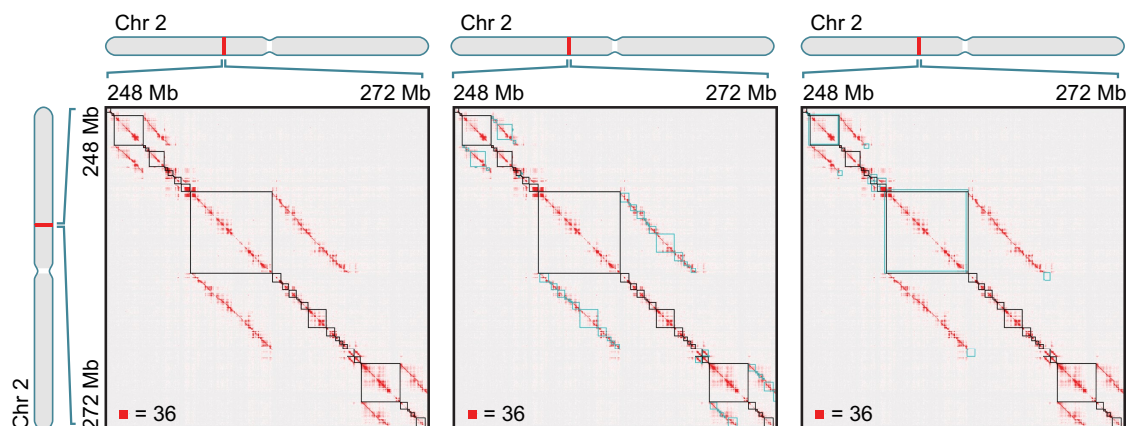
a AGWG project workflow



b Genome size measured by flow cytometry

Species / Strain	Sex	N	Average genome size (Mb)		Statistical analysis
<i>Aedes mascarensis</i>	F	6	1,254	1,254	a
	M	8	1,255		
<i>Aedes aegypti</i> Ho Chi Minh City F13	F	5	1,233	1,231	b
	M	6	1,228		
<i>Aedes aegypti</i> Rockefeller	F	7	1,242	1,227	bc
	M	6	1,213		
<i>Aedes aegypti</i> LVP_AGWG	F	5	1,226	1,224	bc
	M	5	1,222		
<i>Aedes aegypti</i> New Orleans F8	F	8	1,219	1,215	c
	M	7	1,211		
<i>Aedes aegypti</i> Uganda 49-ib-G5	F	5	1,190	1,190	d
	M	6	1,190		

c



d

Chromosome	Assembly	Total size of contigs
Chr 1	Aaegl5, before alternative haplotype removal	542,541,438
	Aaegl5, after alternative haplotype removal	309,614,593 (57%)
	Aaegl4	299,394,366
Chr 2	Aaegl5, before alternative haplotype removal	750,862,705
	Aaegl5, after alternative haplotype removal	473,283,875 (63%)
	Aaegl4	460,653,950
Chr 3	Aaegl5, before alternative haplotype removal	676,921,987
	Aaegl5, after alternative haplotype removal	408,734,344 (60%)
	Aaegl4	397,913,076

e

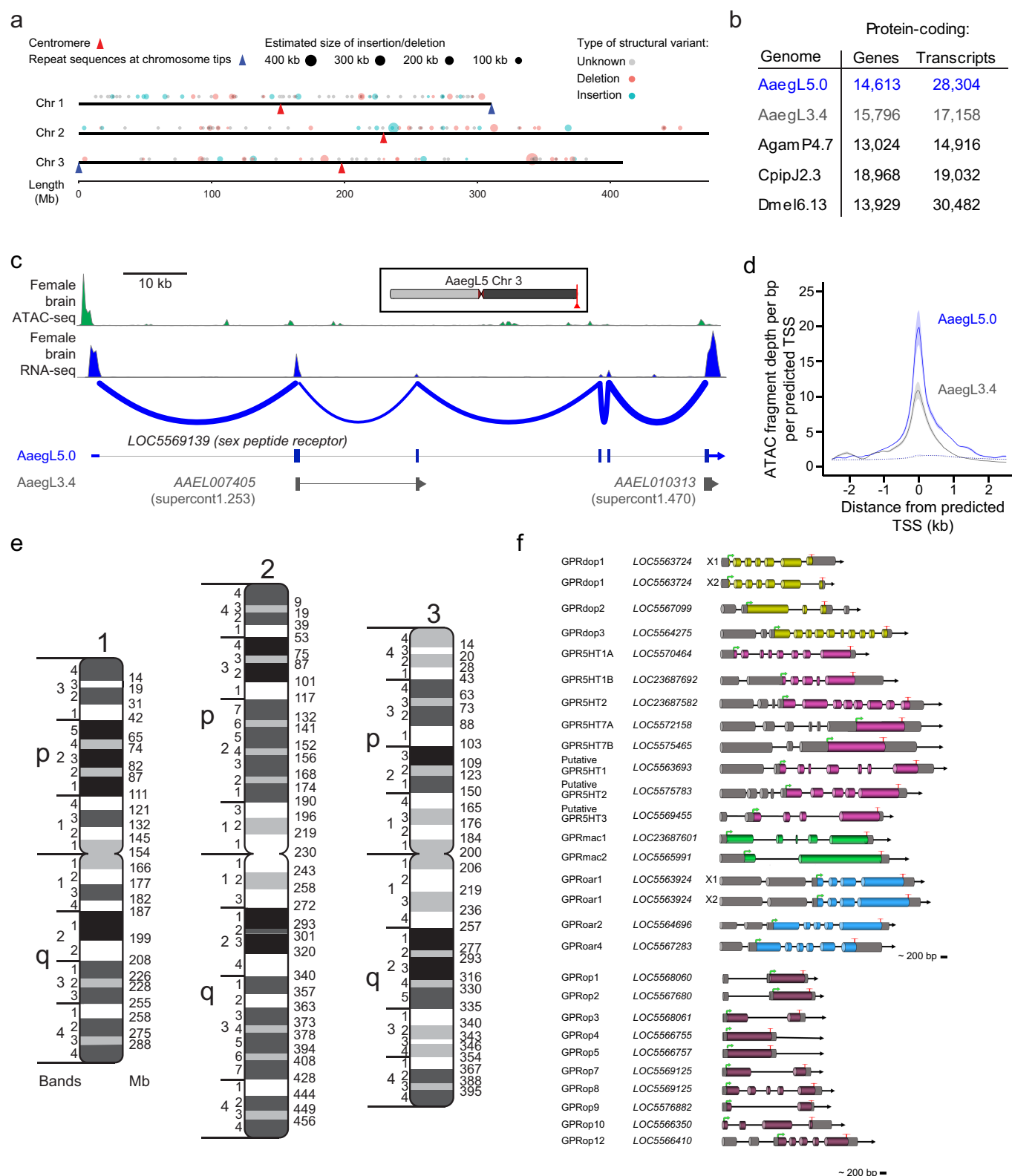
	FALCON-Unzip				Hi-C			
	primary	haplotigs	primary + haplotigs	primary + haplotigs after Hi-C based misjoin correction	scaffolding	alternative haplotype merging	chromosome length scaffolds	small/tiny scaffolds **
Total Sequenced bases	1,695,064,654	351,566,101	2,046,630,755	2,046,630,755	2,046,630,755	1,267,557,260*	1,191,632,812	75,924,448
Number of contigs/gaps	3,967	3,823	7,790	8,306	8,306	2,866	421	2,445
Contig N50	1,304,397	193,091	958,855	929,392	929,392	4,997,917	5,551,291	35,047
Contig NG50 (genome 1.28 Gb)	1,907,936	N/A	1,919,877	1,828,401	1,828,401	4,562,054	4,562,054	N/A
Longest Contig	26,514,109	2,219,489	26,514,109	26,514,109	26,514,109	27,646,994	27,646,994	386,225
Scaffold N50: ***	N/A	N/A	N/A	N/A	677,720,487	408,806,344	408,806,344	36,325

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Project flowchart, measured genome size and assembly process. **a**, Flowchart of LVP_AGWG strain inbreeding, data collection and experimental design of the AegL5 assembly process.

b, Estimated average 1C genome size for each strain for five *Ae. aegypti* strains and *Ae. mascarensis*, the sister taxon of *Ae. aegypti*, for which the genome size has not previously been measured. There were no significant differences between the sexes within and between the species and strains analysed ($P > 0.2$). Significant differences between strains were determined using Proc GLM in SAS with both a Tukey and a Scheffé option with the same outcome. Data labelled with different letters are significantly different ($P < 0.01$). **c**, Combining Hi-C maps with 2D annotations enabled efficient review of sequences identified as alternative haplotypes by sequence alignment. The figure depicts a roughly 24 Mb \times 24 Mb fragment of a contact map generated by aligning a Hi-C dataset to an intermediate genome assembly generated during the process of creating AegL5. This intermediate assembly was a sequence comprising error-corrected, ordered and oriented FALCON-Unzip contigs. The intensity of each pixel in the contact map correlates with how often pairs of loci co-locate in the nucleus. Maximum intensity is indicated in the lower left of each panel. These maps include reads that do not align uniquely (reads with zero mapping quality); such alignments are randomly assigned to one of the possible genomic locations. Three panels show three types of annotations that are overlaid on top of the contact map. Left, FALCON-Unzip contig boundaries are highlighted as black squares along the diagonal. Notably, large linear features appear above and below the diagonal. These are the result of sequence overlap among contigs, which can indicate the presence of undercollapsed heterozygosity in the contig set. Because reads that do not map uniquely are randomly assigned during the alignment step, Hi-C reads derived from a contig will sometimes be

aligned to an overlapping contig. When this happens, the Hi-C read pair may contribute to the formation of a linear feature above and below the diagonal. Therefore, the linear stretches of enriched contact frequency parallel to the diagonal are brought about by the random assignment procedure, and can facilitate the detection of pairs of overlapping contigs. Note that, when the overlap between contigs is owing to undercollapsed heterozygosity, both contigs will exhibit similar long-range contact patterns. This aspect of Hi-C data also provides evidence for the presence of undercollapsed heterozygosity. Centre, LASTZ-alignment-based annotations for fully redundant contigs. The squares shown in blue are obtained by taking diagonal contig boundary annotations (in black) and shifting them up (respectively, left) when drawing above (respectively, below) the diagonal so that the overlapping sequences are horizontally (respectively, vertically) aligned. Note that, as expected, the squares typically span linear, off-diagonal features in the Hi-C data. When one contig is entirely contained in another contig, the redundant contig does not contribute sequence to the merged chromosome-length scaffolds. Right, LASTZ-alignment-based annotations for partially redundant contigs. Again, the squares shown in blue are obtained by taking diagonal contig boundary annotations (in black) and shifting them up and left. The overlaps shown in this panel correspond to contigs that only partially overlap in sequence with other contigs. Consequently, some of their sequence is incorporated in the final fasta. **d**, Comparison of chromosome lengths between AegL4 and AegL5. Numbers are given before post-Hi-C polishing and gap closing. **e**, Step-wise assembly statistics for Hi-C scaffolding, alternative haplotype removal and annotation. *Removed length, 779,073,495 bp. **The definition of scaffold groups can be found in a previously published study⁴. ***Gaps between contigs were set to 500 bp for calculating scaffold statistics. N/A, not applicable.

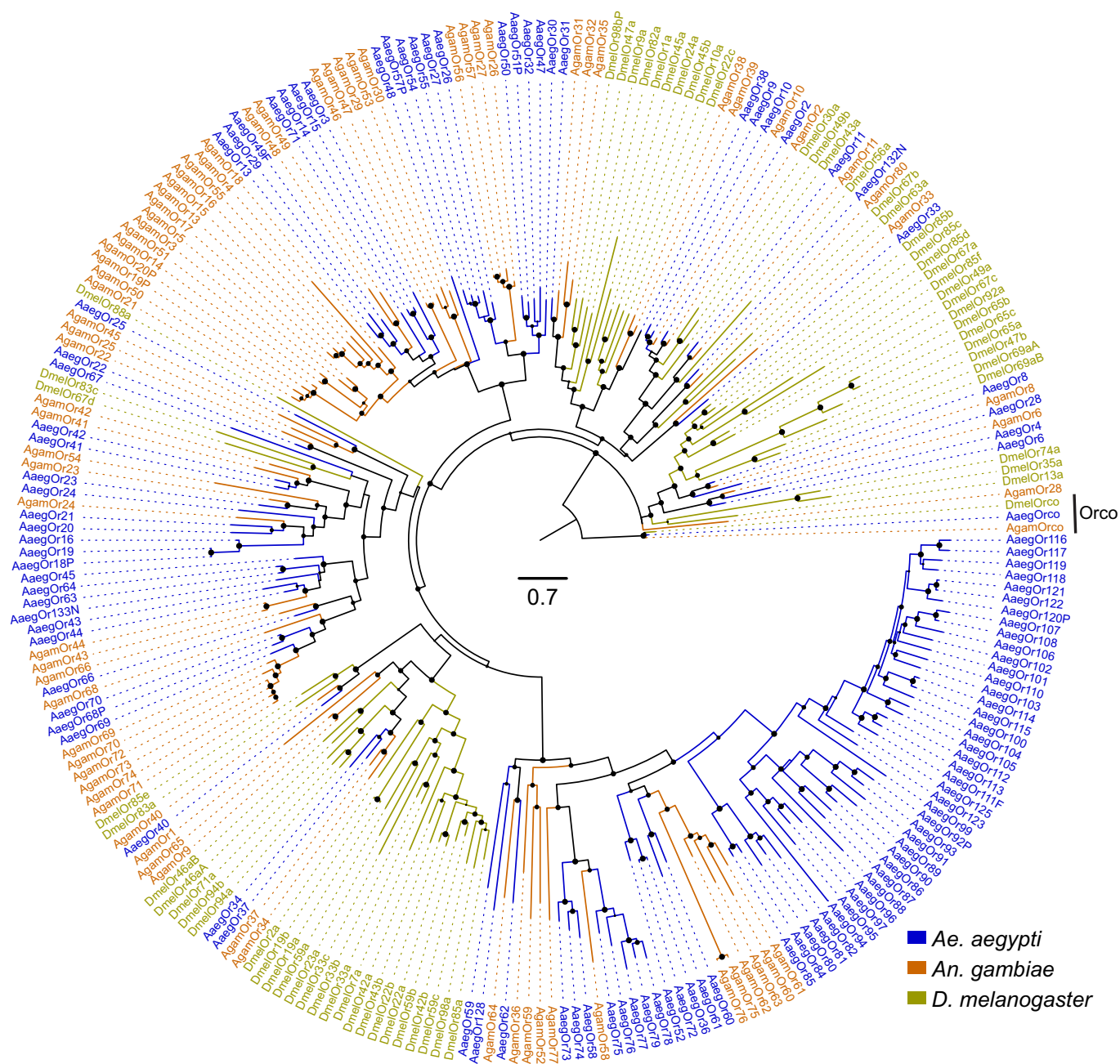


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Remaining assembly gaps, summary of geneset annotation improvement, chromatin accessibility analysis, physical genome map and gene structures of biogenic amine-binding receptors and opsins in AaegL5. **a**, Representation of structural variants identified at assembly gaps by alignment of Bionano optical maps. The estimated size of an insertion (blue) or deletion (red) relative to the reference is represented by the size of the circle. When the size or type of structural variants could not be determined or did not agree between the two optical maps, the location of the assembly gap is plotted in grey. Approximate locations of the centromeres (red triangles) and telomere-associated repeat sequences (blue triangles) are indicated. Raw data are available as Supplementary Data 1. **b**, Comparison of protein-coding genes and transcripts in AaegL5.0 (NCBI RefSeq Release 101) and gene set annotations from *An. gambiae* (Agam), *Culex pipiens* (Cpip) and *D. melanogaster* (Dmel). **c**, *Sex peptide receptor* structure in AaegL3.4 and AaegL5.0, and female brain RNA-seq and ATAC-seq reads aligned to AaegL5. Blue lines on the RNA-seq track indicate splice junctions, with the number of reads spanning a junction represented by line thickness. Exons are represented by tall filled boxes and introns by lines. Arrowheads

indicate gene orientation. **d**, Average read profiles across promoter regions, defined as the transcription start site (TSS) ± 2.5 kb. Solid lines represent Tn5-treated native chromatin using the ATAC-seq protocol ($n = 4$), dotted lines represent Tn5-treated naked genomic DNA ($n = 1$). Shaded regions represent s.d. **e**, A physical genome map was developed by localizing 500 BAC clones to chromosomes using FISH. For the development of a final chromosome map for the AaegL5 assembly, we assigned the coordinates of each outmost BAC clone within a band (Supplementary Data 12) to the boundaries between bands. The final resolution of this map varies on average between 5 and 10 Mb because of the differences in BAC mapping density in different regions of chromosomes. **f**, Schematic of predicted gene structures of the *Ae. aegypti* biogenic amine-binding receptors and opsins. Exons, cylindrical bars; introns, black lines; dopamine receptors, yellow bars; serotonin receptors, magenta bars; muscarinic acetylcholine receptors, green bars; octopamine receptors, blue bars; opsins, dark purple bars; predicted 3' and 5' non-coding sequence (dark shading). The 'unclassified receptor' *GPRnna19* is not shown. Details on gene models compared to previous annotations and the predicted amino acid sequences of each gene are available in Supplementary Data 14–16.

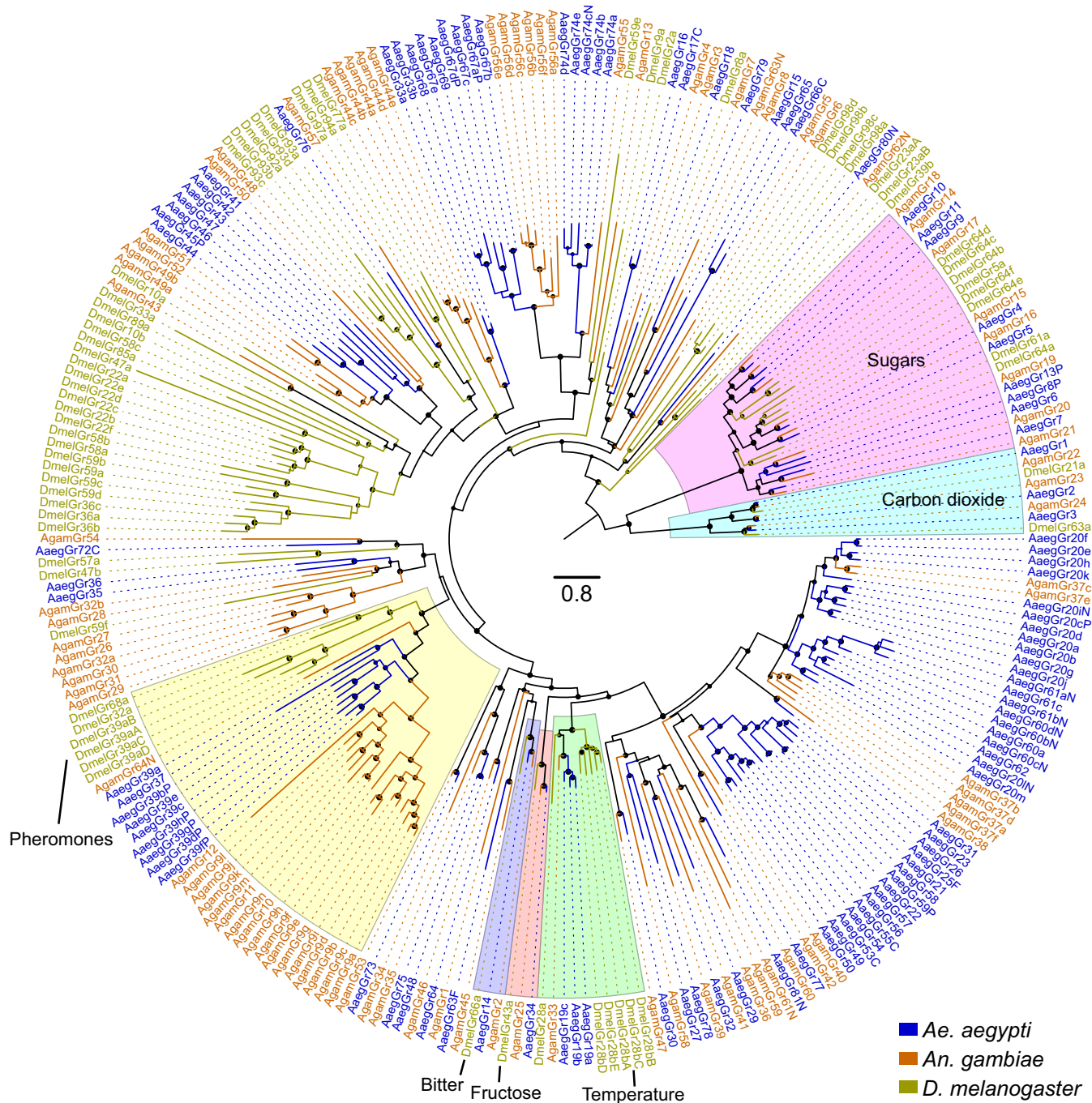
Odorant receptors (ORs)



Extended Data Fig. 4 | Phylogenetic tree of odorant receptor gene families from *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. Maximum likelihood odorant receptor tree was rooted with Orco proteins, which are both highly conserved and basal within the odorant receptor family⁸⁹. Support levels for nodes are indicated by the size of black circles—

reflecting approximate likelihood ratio tests (aLRT values ranging from 0 to 1 from PhyML v.3.0 run with default parameters⁹⁰). Suffixes after protein names are C, minor assembly correction; F, major assembly modification; N, new model; P, pseudogene. Scale bar, amino acid substitutions per site.

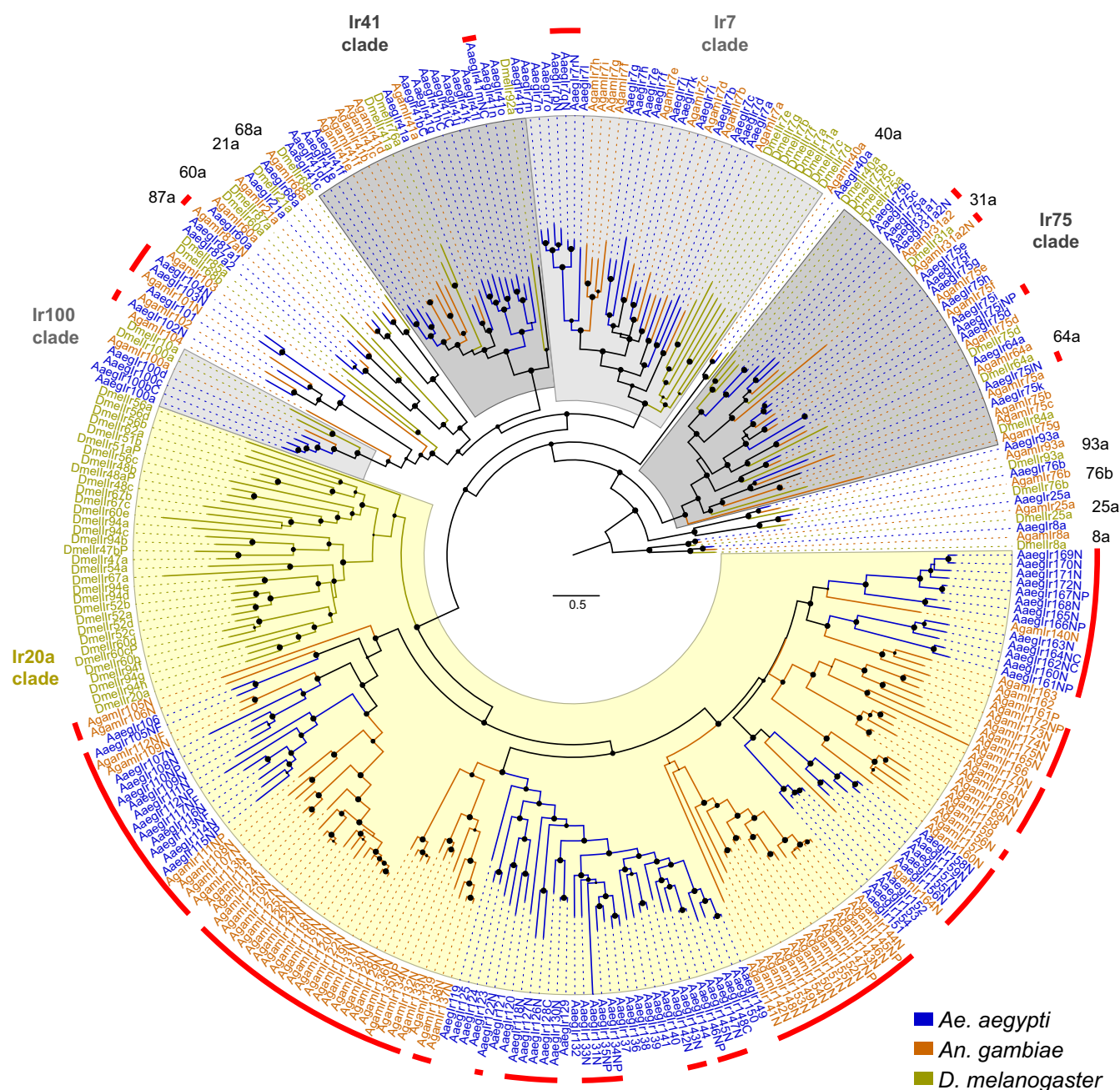
Gustatory receptors (GRs)



Extended Data Fig. 5 | Phylogenetic tree of the gustatory receptor gene families from *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. Maximum likelihood gustatory receptor tree was rooted with the highly conserved and distantly related carbon dioxide and sugar receptor subfamilies, which together form a basal clade within the arthropod gustatory receptor family⁸⁹. Subfamilies and lineages closely related to *D. melanogaster*

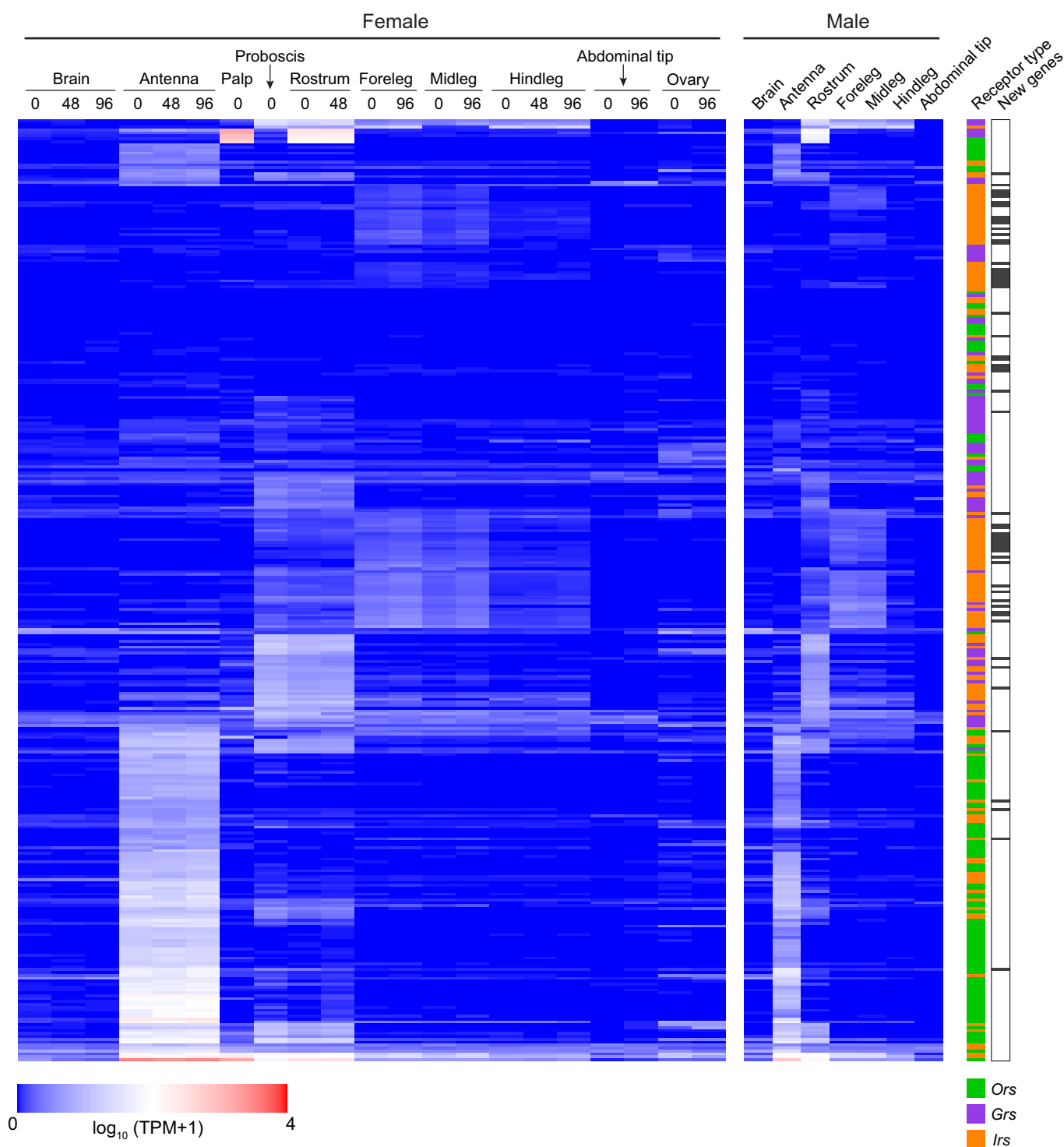
gustatory receptors of known function are highlighted. Support levels for nodes are indicated by the size of black circles—reflecting approximate likelihood ratio tests (aLRT values ranging from 0 to 1 from PhyML v.3.0 run with default parameters⁹⁰). Suffixes after protein names are C, minor assembly correction; F, major assembly modification; N, new model; P, pseudogene. Scale bar, amino acid substitutions per site.

Ionotropic receptors (IRs)



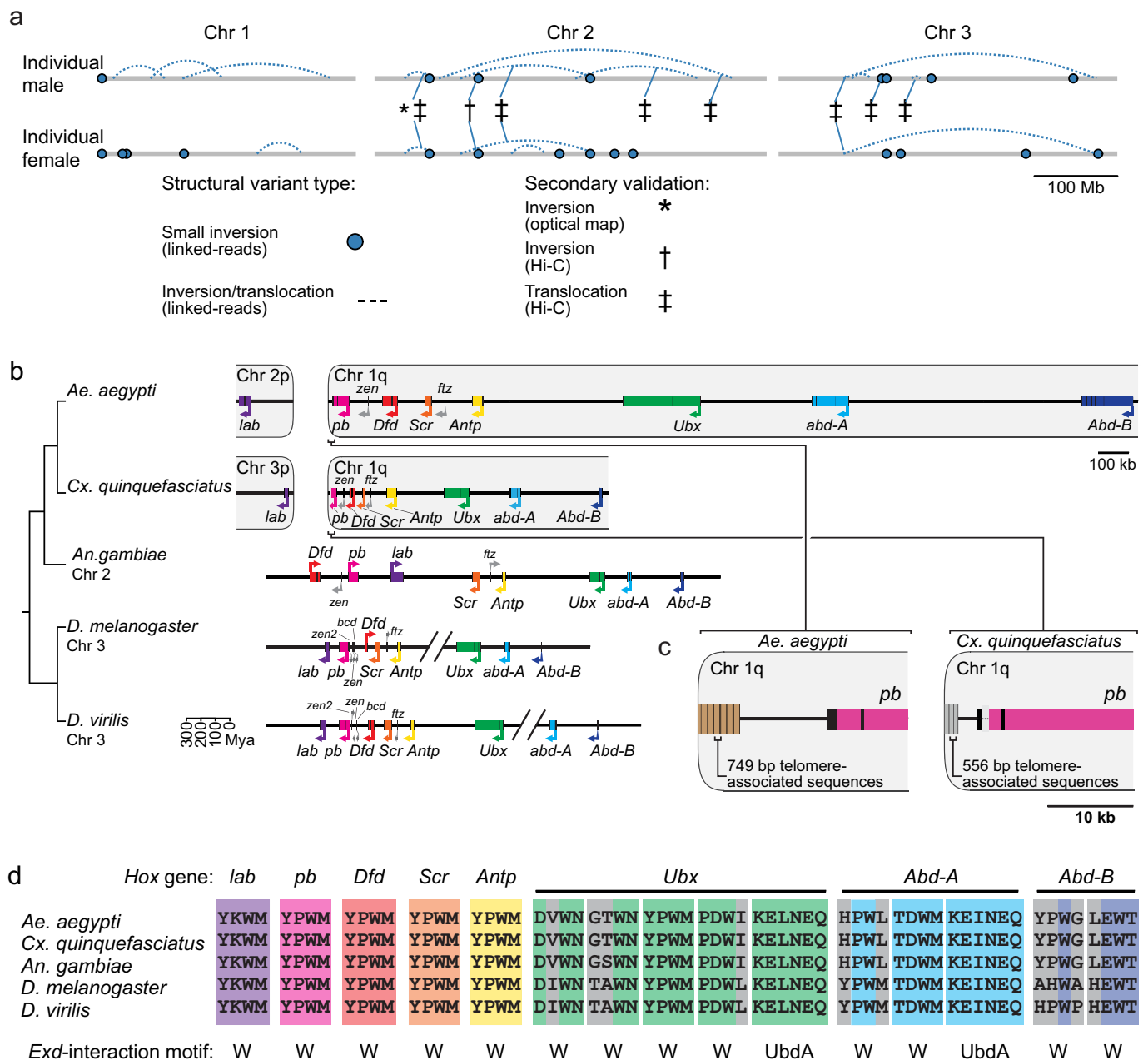
Extended Data Fig. 6 | Phylogenetic tree of the ionotropic receptor gene families from *Ae. aegypti*, *An. gambiae* and *D. melanogaster*. Maximum likelihood phylogenetic tree of ionotropic receptor protein sequences from the indicated species rooted with highly conserved Ir8a and Ir25a proteins. Conserved proteins with orthologues in all species are named outside the circle, and previously unannotated ionotropic receptors are highlighted

with red lines. Support levels for nodes are indicated by the size of black circles—reflecting approximate likelihood ratio tests (aLRT values ranging from - to 1 from PhyML v.3.0 run with default parameters⁹⁰). Suffixes after protein names are C, minor assembly correction; F, major assembly modification; N, new model; P, pseudogene. Scale bar, amino acid substitutions per site.



Extended Data Fig. 7 | Chemosensory receptor expression in adult *Ae. aegypti* tissues. Previously published RNA-seq data¹³ were reanalysed using the new chemoreceptor annotations and genome assembly. Chemosensors have been clustered according to Euclidian distance of their expression vectors using the R function `hclust`. Expression is given

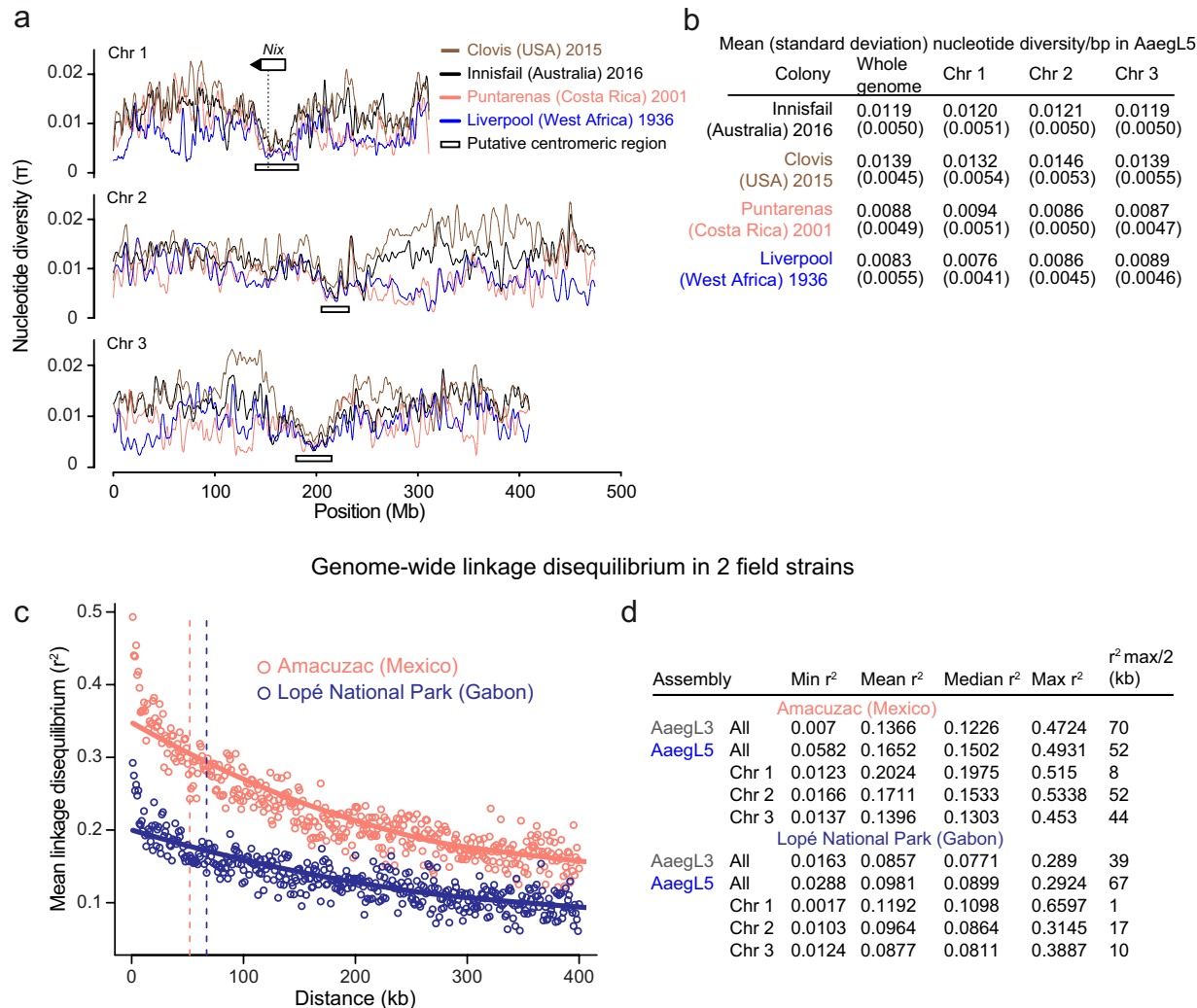
for females at three stages of the gonotrophic cycle (0, 48 or 96 h after taking a blood-meal, for which 0 h indicates not blood-fed, 48 h indicates 48 h after the blood-meal, and 96 h indicates gravid). New genes are indicated by black bars on the right.



Extended Data Fig. 8 | Structural variation, the Hox gene cluster and Hox cofactor motifs. **a**, Linked-read sequencing of two individuals from the LVP_AGWG strain identified putative structural variants in the AagL5 assembly. **b**, Comparative genomic arrangement of the Hox cluster (HOXC) in five species (Supplementary Data 22). Note the split of *labial* (*lab*) and *proboscipedia* (*pb*) between two chromosomes in *Ae. aegypti* and *Cx. quinquefasciatus*. Owing to chromosome arm

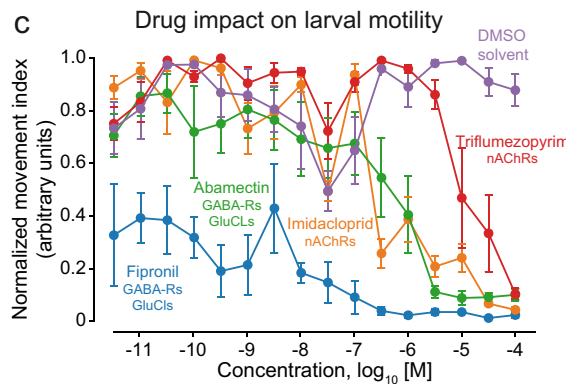
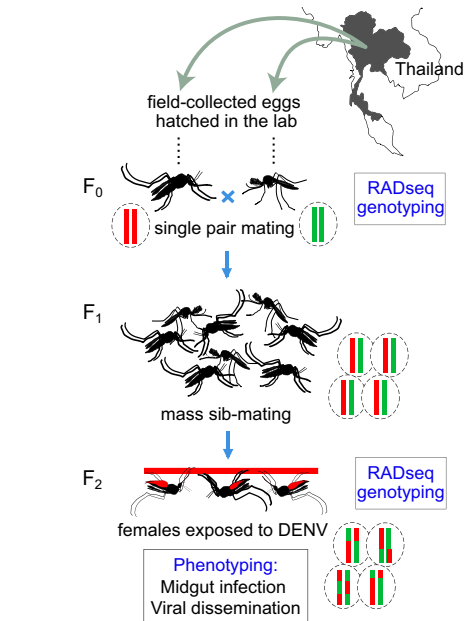
exchange, chromosome 3p in *Cx. quinquefasciatus* is the homologue of chromosome 2p in *Ae. aegypti*⁴. **c**, Repeats in putative telomere-associated sequences downstream of *pb* in both species. **d**, Motifs known to mediate protein–protein interactions with the Hox cofactor Extradenticle (*Exd*)⁹¹ from the five indicated species are aligned using Clustal-Omega. Perfectly aligned residues are coloured according to Hox gene identity, non-conserved residues are grey.

Genome-wide genetic variation in 4 colonized strains



Extended Data Fig. 9 | Population genomic structure and linkage disequilibrium analysis of *Ae. aegypti* strains. **a**, Chromosomal patterns of nucleotide diversity (π) in four strains of *Ae. aegypti* measured in 100-kb non-overlapping windows and presented as a LOESS-smoothed curve. **b**, Mean nucleotide diversity in the strains in **a**, with s.d. indicated in parentheses. Nucleotide diversity (π) was measured in non-overlapping 100-kb windows. The Liverpool and Costa Rica colonies maintain extensive diversity despite being colonized in the laboratory more than a decade ago, but show reduced genome-wide diversity (on the order of 30–40%) relative to the more recently laboratory colonized Innisfail

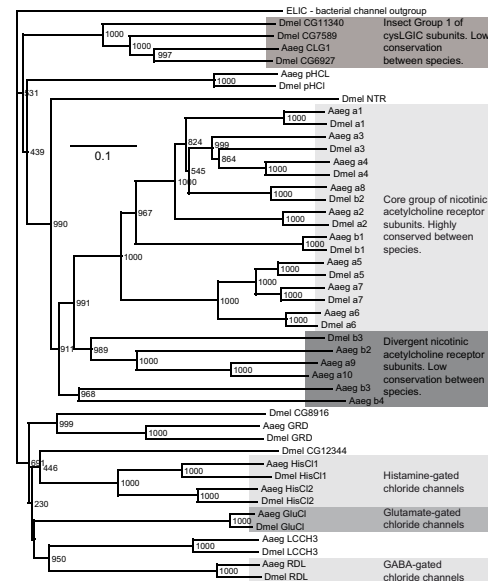
and Clovis. **c**, Pairwise linkage disequilibrium between SNPs located within the same chromosome estimated from 28 wild-caught individuals from the indicated populations. Each point represents the mean linkage disequilibrium for that set of binned SNP pairs. Solid lines are LOESS-smoothed curves, and dashed lines correspond to $r^2_{\max}/2$. Inclusion of additional individuals available from the Amacuzac population (up to 137) had a minimal effect on the linkage disequilibrium estimations ($\Delta R^2 < 0.017$; data not shown). **d**, Linkage disequilibrium (r^2) values along the *Ae. aegypti* AaegL5 genome assembly based on pairwise SNP comparisons. Data were obtained from the average r^2 of SNPs in 1-kb bins.

a Experimental design, DENV susceptibility

Extended Data Fig. 10 | QTL analysis of DENV competence in *Ae. aegypti* and Cys-loop LGICs. **a**, Schematic representation of the experimental workflow for testing DENV competence in *Ae. aegypti*, related to Fig. 5b–d. **b**, Comparison of QTL map density constructed against AaegL3 or AaegL5 assemblies. **c**, Concentration–response curves showing the effect on *Ae. aegypti* larval motility of insecticides currently used in veterinary and agricultural applications (mean \pm s.e.m., $n = 7$). **d**, Phylogenetic tree of Cys-loop LGIC subunits for *Ae. aegypti* and *D. melanogaster*. The accession numbers of the *D. melanogaster* sequences used in constructing the tree are: D α 1 (CAA30172), D α 2

b QTL comparison (AaegL3 vs. AaegL5)

	Chr1	Chr2	Chr3	Overall
AaegL5-guided map				
Number of markers mapped	76	80	99	255
Maximum marker spacing (cM)	16.8	12	11.5	16.8
Average marker spacing (cM)	2.1	2.3	1.1	1.8
Length of linkage group (cM)	159.6	183.1	106	448.7
AaegL3-guided map (restricted to sex-linked region for Chr. 1)				
Number of markers mapped	12	32	33	77
Maximum marker spacing (cM)	10.0	17.5	8.6	17.5
Average marker spacing (cM)	3.3	2.2	2.1	2.3
Length of linkage group (cM)	36.8	67.8	66.5	171.2

d Ligand-gated ion channels (LGICs)

(CAA36517), D α 3 (CAA75688), D α 4 (CAB77445), D α 5 (AAM13390), D α 6 (AAM13392), D α 7 (AAK67257), D β 1 (CAA27641), D β 2 (CAA39211), D β 3 (CAC48166), GluCl (AAG40735), GRD (Q24352), HisCl1 (AAL74413), HisCl2 (AAL74414), LCCH3 (AAB27090), Ntr (NP_651958), pHCl (NP_001034025), RDL (AAA28556). For *Ae. aegypti* sequences, see Supplementary Data 24. ELIC (Erwinia ligand-gated ion channel), which is an ancestral Cys-loop LGIC found in bacteria (accession number P0C7B7), was used as an outgroup. Scale bar, amino acid substitutions per site.

TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle

Thomas O. Vogler^{1,2,15}, Joshua R. Wheeler^{2,3,15}, Eric D. Nguyen^{2,4}, Michael P. Hughes^{5,6}, Kyla A. Britson^{7,8}, Evan Lester^{2,3}, Bhalchandra Rao³, Nicole Dalla Betta¹, Oscar N. Whitney¹, Theodore E. Ewachiw¹, Edward Gomes⁹, James Shorter⁹, Thomas E. Lloyd^{7,8}, David S. Eisenberg^{5,6,10}, J. Paul Taylor^{11,12}, Aaron M. Johnson^{4,13}, Bradley B. Olwin^{1*} & Roy Parker^{3,14*}

A dominant histopathological feature in neuromuscular diseases, including amyotrophic lateral sclerosis and inclusion body myopathy, is cytoplasmic aggregation of the RNA-binding protein TDP-43. Although rare mutations in *TARDBP*—the gene that encodes TDP-43—that lead to protein misfolding often cause protein aggregation, most patients do not have any mutations in *TARDBP*. Therefore, aggregates of wild-type TDP-43 arise in most patients by an unknown mechanism. Here we show that TDP-43 is an essential protein for normal skeletal muscle formation that unexpectedly forms cytoplasmic, amyloid-like oligomeric assemblies, which we call myo-granules, during regeneration of skeletal muscle in mice and humans. Myo-granules bind to mRNAs that encode sarcomeric proteins and are cleared as myofibres mature. Although myo-granules occur during normal skeletal-muscle regeneration, myo-granules can seed TDP-43 amyloid fibrils in vitro and are increased in a mouse model of inclusion body myopathy. Therefore, increased assembly or decreased clearance of functionally normal myo-granules could be the source of cytoplasmic TDP-43 aggregates that commonly occur in neuromuscular disease.

The function and aggregation of the RNA-binding protein TAR DNA-binding protein 43 (TDP-43) in multinucleated skeletal-muscle cells (myofibres) is of interest for two reasons. First, TDP-43 aggregates accumulate in the skeletal muscle of patients with inclusion body myopathy, oculopharyngeal muscular dystrophy and distal myopathies^{1,2}. These aggregates appear to be similar to the cytoplasmic TDP-43 aggregates that are found in the neurons of patients with amyotrophic lateral sclerosis and frontotemporal lobar degeneration, suggesting that there is a common mechanism in muscle and neurons that leads to histopathological, cytoplasmic TDP-43 aggregation^{2–4}. Second, reducing the levels of TDP-43 leads to age-related muscle weakness in mice⁵, muscle degeneration and sarcomere disruption in zebrafish⁶ and age-related muscle weakness in *Drosophila* wing muscles^{7,8}. Given the requirement for TDP-43 in muscle function and its potential to form cytoplasmic aggregates in muscle diseases, we examined TDP-43 function during normal mammalian skeletal-muscle formation.

Cytoplasmic TDP-43 myo-granules

We first examined the subcellular distribution of TDP-43 in cultured skeletal-muscle cells and found abundant nuclear TDP-43 in C2C12 myoblasts, a mouse muscle cell line⁹ (Extended Data Fig. 1a). However, during the differentiation of C2C12 myoblasts and isolated primary mouse myoblasts into multinucleated myotubes, we observed an increase in cytoplasmic TDP-43 by immunofluorescence and by subcellular fractionation (Extended Data Fig. 1a–e). Furthermore, live-cell single-molecule imaging of HaloTag–TDP-43 revealed increased cytosolic HaloTag–TDP-43 in differentiating myotubes compared to myoblasts (Extended Data Fig. 1f–k and Supplementary Videos 1, 2).

We next examined the subcellular distribution of TDP-43 in uninjured tibialis anterior muscle and tibialis anterior muscle that was chemically injured using barium chloride (BaCl₂) and was subsequently allowed to regenerate^{10,11} (Fig. 1a). Although we primarily observed nuclear TDP-43 in uninjured muscle, at five days post-injury (DPI) TDP-43 was upregulated and was found to be expressed in both the myonuclei and cytoplasm of newly forming myofibres, identified by embryonic myosin heavy chain (eMHC) expression (Fig. 1b and Extended Data Fig. 2a). Super-resolution microscopy revealed that TDP-43 was predominately localized to myonuclei in uninjured myofibres, whereas in regenerating myofibres cytoplasmic TDP-43 localized to regions surrounding eMHC (Fig. 1c), which are sites of newly formed sarcomeres¹². By 10 DPI, cytosolic TDP-43 levels decreased and TDP-43 localized around the centrally located nuclei, whereas by 30 DPI, TDP-43 had relocated to the nucleus (Fig. 1b, d and Extended Data Fig. 2b). Therefore, cytosolic TDP-43 increases during skeletal-muscle-cell formation in culture and in mice.

Because cytoplasmic TDP-43 localization is associated with pathological aggregation, we investigated whether the cytoplasmic TDP-43 identified in skeletal-muscle formation adopts a higher-order, oligomeric state. We detected an increase in urea-insoluble TDP-43 in myotubes compared to myoblasts and observed higher molecular weight SDS-resistant TDP-43 assemblies unique to differentiating myotubes by semi-denaturing detergent agarose-gel electrophoresis (Extended Data Fig. 2c, d). Furthermore, immunoprecipitation of TDP-43 from myotubes and from mouse tibialis anterior muscles at 5 DPI, followed by transmission electron microscopy (TEM), reveals the

¹Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO, USA. ²Medical Scientist Training Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ³Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA. ⁴Molecular Biology Program, Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ⁵Department of Biological Chemistry, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ⁶Department of Chemistry and Biochemistry, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ⁷Departments of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁸Departments of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁹Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ¹⁰Howard Hughes Medical Institute, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. ¹¹Department of Cell and Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN, USA. ¹²Howard Hughes Medical Institute, St. Jude Children's Research Hospital, Memphis, TN, USA. ¹³University of Colorado School of Medicine RNA Bioscience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ¹⁴Howard Hughes Medical Institute, University of Colorado, Boulder, CO, USA. ¹⁵These authors contributed equally: Thomas O. Vogler, Joshua R. Wheeler. *e-mail: olwin@colorado.edu, roy.parker@colorado.edu

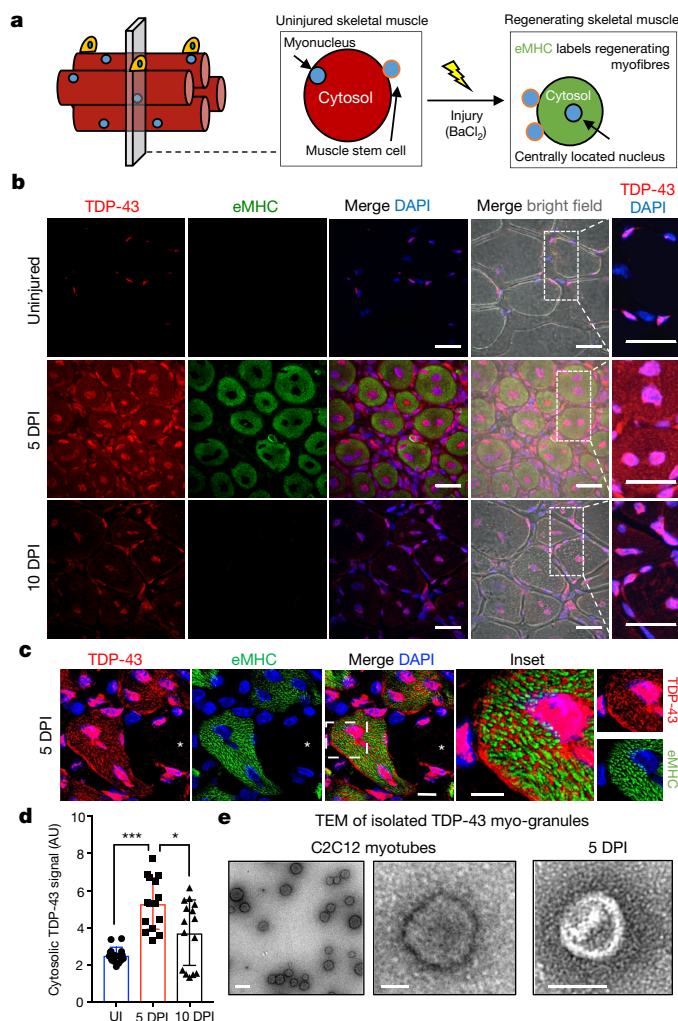


Fig. 1 | TDP-43 adopts a higher-order state during normal skeletal-muscle formation. **a**, Schematic of regeneration of skeletal muscle injuries in wild-type mice. **b**, TDP-43 expression after BaCl_2 -induced tibialis anterior muscle injury. eMHC expression in regenerating myofibres; nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI). Scale bars, 25 μm . $n = 5$ mice per condition, each showing similar results. **c**, Super-resolution imaging of TDP-43 expression around nascent sarcomeres in the cytoplasm during muscle regeneration. Scale bar, 10 μm and 5 μm (inset). Asterisk, an uninjured myofibre that lacks eMHC and TDP-43 cytosolic signals. Nuclei are counterstained with DAPI. $n = 3$ biologically independent experiments, each showing similar results. **d**, Quantification of the cytoplasmic TDP-43 signal in skeletal muscle myofibres. AU, arbitrary units. Unpaired, two-tailed Student's *t*-tests were used for each individual comparison: 5 DPI versus uninjured (UI), *** $P = 4.36 \times 10^{-8}$; 5 DPI versus 10 DPI, * $P = 0.011$; 10 DPI versus UI, $P = 0.015$. $n = 3$ biological replicates, $n = 5$ myofibres per replicate. Data are mean \pm s.d. **e**, TEM images of myo-granules isolated by TDP-43 immunoprecipitation obtained from C2C12 myotubes and from mouse tibialis anterior muscle at 5 DPI. $n = 3$ biologically independent experiments, each showing similar results.

presence of 50–250-nm assemblies that are not detected in undifferentiated myoblasts or in uninjured skeletal muscle (Fig. 1e and Extended Data Fig. 2e, f). The TEM structure is similar to previously characterized TDP-43 oligomers, albeit roughly twofold larger in diameter¹³. To exclude the possibility that the TDP-43 assemblies in skeletal muscle are stress granules, we assayed C2C12 myotubes for the stress granule markers G3BP1 and PABP1. Stress granules were not present during normal myotube formation (Extended Data Fig. 2g). Therefore, during muscle formation, TDP-43 exists as a component of an SDS-resistant oligomeric assembly that is distinct from stress granules and that we refer to as myo-granules.

Myo-granules are amyloid-like assemblies

The SDS resistance of these myo-granules suggests that myo-granules have amyloid-like properties, which is supported by two observations. First, X-ray diffraction of lyophilized myo-granules revealed a diffraction pattern with a 4.8 Å reflection, indicating a β -rich complex that is not observed in control samples. Myo-granules lacked a 10 Å reflection, which suggests a lack of mated cross β -sheets similar to previously described amyloid-like oligomers¹⁴ (Fig. 2a and Extended Data Fig. 3a, b). Second, immunopurified myo-granules from C2C12 myotubes and regenerating mouse tibialis anterior muscle also show A11 immunoreactivity, a conformation-specific antibody that recognizes β -rich structures, including amyloid-like oligomers¹⁵ (Extended Data Fig. 3c–h).

Similar to TDP-43, A11 immunoreactivity increases in myotubes in culture and in regenerating mouse tibialis anterior muscle (Extended Data Fig. 3i–k). In developing myotubes in culture, A11 immunoreactivity is cytoplasmic and correlates with cytoplasmic TDP-43 expression (Extended Data Fig. 4a–c). During muscle regeneration A11 immunoreactivity correlates with TDP-43 cytoplasmic expression, increasing in the cytoplasm at 5 DPI but disappearing by 10 DPI (Fig. 2b and Extended Data Fig. 4d–g). At 5 DPI in mice, more than 80% of A11 immunoreactivity is co-localized with cytosolic TDP-43 expression (Fig. 2c, d and Extended Data Fig. 4h). Furthermore, cytoplasmic TDP-43 exists as a component of an A11-reactive complex, as revealed by proximity ligation assays in differentiating C2C12 myotubes (Extended Data Fig. 4i). These observations indicate that cytoplasmic myo-granules contain TDP-43 in an amyloid-like oligomer conformation during skeletal muscle formation.

Myo-granules contain sarcomeric mRNAs

Because TDP-43 is an RNA-binding protein, we examined whether myo-granules included RNA. Immunoprecipitation of myo-granules with TDP-43 or A11 antibodies followed by oligo-dT northern blot analysis reveals that TDP-43 and A11 associate with mRNA in myotubes (Extended Data Fig. 5a). To identify the mRNAs that are bound by TDP-43 during muscle formation, we constructed transcriptome-wide maps of TDP-43-binding sites in undifferentiated myoblasts and in myotubes using enhanced ultraviolet-light crosslinking and immunoprecipitation (eCLIP)¹⁶ (Extended Data Fig. 5b–e). We identified a total of 556 binding sites across 174 genes for myoblasts and a total of 975 binding sites across 320 genes for myotubes that were significantly enriched compared to size-matched input (which reflects all RNA–protein interactions in the input). The binding sites were highly correlated between biological replicates, showed enrichment for the TDP-43 UG-rich consensus sequence and had thousands of reproducible CLIP clusters as shown by irreproducible discovery rate analysis; we identified known TDP-43 mRNA targets including the 3' untranslated region of TDP-43¹⁷ (Extended Data Fig. 5f–h). We also observed that the mRNAs that bind to TDP-43 changed significantly during skeletal muscle differentiation (Extended Data Fig. 6a).

Most of the TDP-43 binding sites in myoblasts and myotubes were found to be in exons of protein-coding transcripts, suggesting that TDP-43 could be associating with mature mRNAs (Fig. 3a and Extended Data Fig. 6b, c). By contrast, TDP-43-binding sites in neurons predominantly mapped to introns^{18,19}. The difference may reflect cell state, whereby TDP-43 binds more processed cytoplasmic RNAs in newly forming tissue and more nuclear intronic RNA in post-mitotic mature cells. Connectome and Gene Ontology analysis of TDP-43 exonic target transcripts in myotubes, which are likely to constitute interactions with cytoplasmic mRNAs, revealed that TDP-43 binds to a network of transcripts associated with the sarcomere (Fig. 3b and Extended Data Fig. 6a). TDP-43 target RNAs that were identified by eCLIP in myotubes often had multiple TDP-43 exonic binding sites in close proximity¹⁹. For example, numerous TDP-43 exon-binding sites are distributed across the mRNA transcript of *Ttn* (which encodes titin), and we observed multiple UG-rich

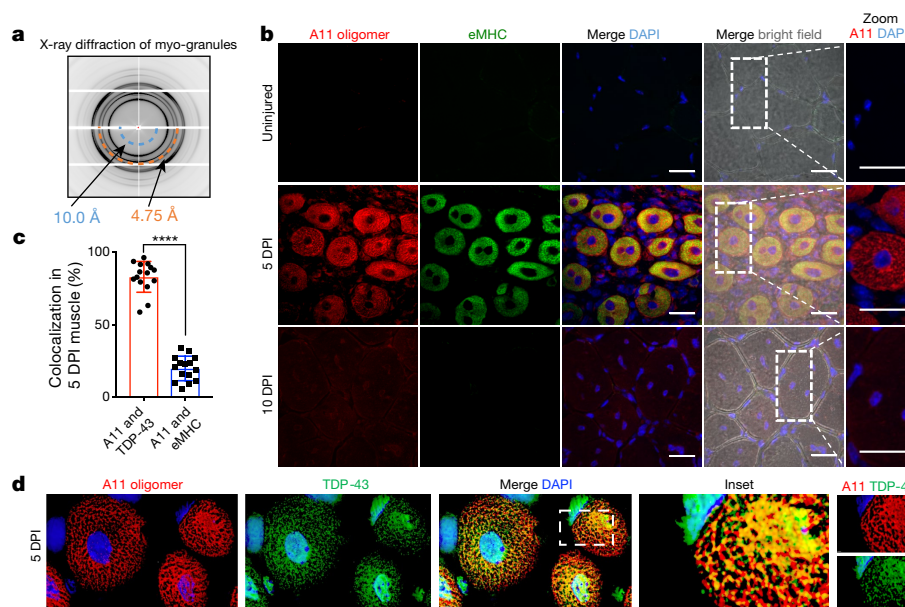


Fig. 2 | Myo-granules containing TDP-43 are amyloid-like oligomers. **a**, X-ray diffraction of myo-granules immunoprecipitated from C2C12 myotubes. Two rings at approximately 4.8 Å (orange) and approximately 10 Å (blue) are drawn on the bottom half to highlight the locations of these reflections. One sample per condition was used. Two diffraction images at different rotations were taken per sample and each image showed similar results. **b**, A11 immunoreactivity in tibialis anterior muscle regeneration and uninjured muscle. $n = 4$ mice per condition. Regenerating myofibres

showed eMHC expression. Scale bars, 25 μm . **c**, Quantification of A11 and TDP-43 co-localization and A11 and eMHC co-localization in skeletal muscle at 5 DPI. Unpaired, two-tailed Student's t -test, **** $P = 6.3 \times 10^{-17}$, $n = 3$ mice, $n = 5$ myofibres per mouse. Data are mean \pm s.d. **d**, Representative deconvolution images of A11 and TDP-43 co-localization in mouse tibialis anterior myofibres at 5 DPI for data quantified in **c**. Scale bars, 3 μm and 1 μm (inset). $n = 3$ mice, each showing similar results.

stretches within single exons of *Ttn* with several TDP-43-binding sites (Extended Data Fig. 6d). These observations suggest that TDP-43 has a different function during myogenesis, during which TDP-43 binds to structural mRNAs that are required for skeletal muscle formation, while retaining canonical nuclear functions such as splicing and nuclear cytoplasmic shuttling.

To confirm that the sarcomeric mRNAs that were identified by eCLIP bind to cytoplasmic TDP-43 during muscle formation, we used single-molecule fluorescence in situ hybridization. We found that the TDP-43 protein co-localizes with mRNAs of *Myh3* (which encodes eMHC) and *Tnni1* in the cytoplasm of myotubes (Fig. 3c). In addition, single-molecule fluorescence in situ hybridization of *Ttn* mRNA reveals co-localization of both TDP-43 expression and A11 immunoreactivity with *Ttn* mRNA in myotubes (Fig. 3d). These observations demonstrate that TDP-43 binds to sarcomeric mRNAs in the cytosol, and can form A11-positive myo-granules in association with those mRNAs—perhaps because of the high local concentration of TDP-43 proteins on a single mRNA molecule^{19,20}.

The association of TDP-43 myo-granules with sarcomeric mRNAs during muscle formation is analogous to the role of TDP-43 in forming cytoplasmic neuronal messenger ribonucleoprotein granules for local translation of mRNAs in neurons²¹. Consistent with this similarity, mass spectrometry of purified myo-granules identified 356 proteins that were enriched in proteins involved in RNA localization and translation, which overlaps with the TDP-43 interactome²² and the neuronal RNA granule proteome²³ (Extended Data Fig. 7a–d and Supplementary Tables 1, 2). Myo-granules included valosin-containing protein (VCP), a protein that has been linked to neuromuscular degeneration²⁴; we validated this by analysing the co-localization of VCP with A11 and TDP-43 in the cytoplasm of regenerating muscle (Extended Data Fig. 7e). However, HNRNPA2B1, an RNA-binding protein associated with neuromuscular degeneration²⁵, was not identified in myo-granules and remained in the nucleus during muscle regeneration (Extended Data Fig. 7f). Therefore, myo-granules associate with a specific set of proteins that may help to localize and regulate sarcomeric mRNAs during skeletal-muscle formation.

TDP-43 is essential for muscle formation

If TDP-43-containing myo-granules are sarcomeric messenger ribonucleoproteins, then genetic depletion of *Tardbp* may disrupt skeletal-muscle myofibre formation. CRISPR-Cas9-mediated deletion of *Tardbp* in C2C12 cells arrested growth of C2C12 myoblasts, which led to cell death and prevented myoblast differentiation (Fig. 3e, f and Extended Data Fig. 8a). Because TDP-43 appears to be essential for myoblast proliferation and survival, we investigated whether removing one copy of the *Tardbp* gene in muscle stem cells using *Pax7*^{iresCre} recombination impaired muscle regeneration^{26,27} (Extended Data Fig. 8b, c). In uninjured mice, the size of myofibres and number of muscle stem cells were unaffected when one copy of *Tardbp* was deleted from muscle stem cells (Extended Data Fig. 8d–f). However, after injury, mice with one *Tardbp* allele in muscle stem cells had significantly smaller myofibres than wild-type mice (Fig. 3g, h and Extended Data Fig. 8g–i). Because there was no detectable change in the number of muscle stem cells when one *Tardbp* allele was deleted, we hypothesize that the regeneration defect is in part because of loss of TDP-43 function during myofibre formation. Therefore, TDP-43 is essential for skeletal-muscle-cell differentiation in culture and required for skeletal-muscle regeneration.

Myo-granules in humans and disease

To determine whether cytoplasmic TDP-43 and myo-granules are conserved in human muscle regeneration, we examined human muscle biopsies from patients with different clinical and pathological features of necrotizing myopathy. For each patient, we observed increased cytoplasmic TDP-43 and A11 amyloid oligomer staining in the regenerating muscle, indicating that myo-granules form in regenerating human myofibres and are not present in non-regenerating myofibres (Fig. 4 and Extended Data Fig. 9a). It is possible that myo-granules formed during normal regeneration may seed the aggregates seen in human muscle diseases.

Because myo-granules containing TDP-43 form during human skeletal-muscle regeneration and TDP-43 aggregates are found in skeletal-muscle diseases, the increased regeneration occurring in

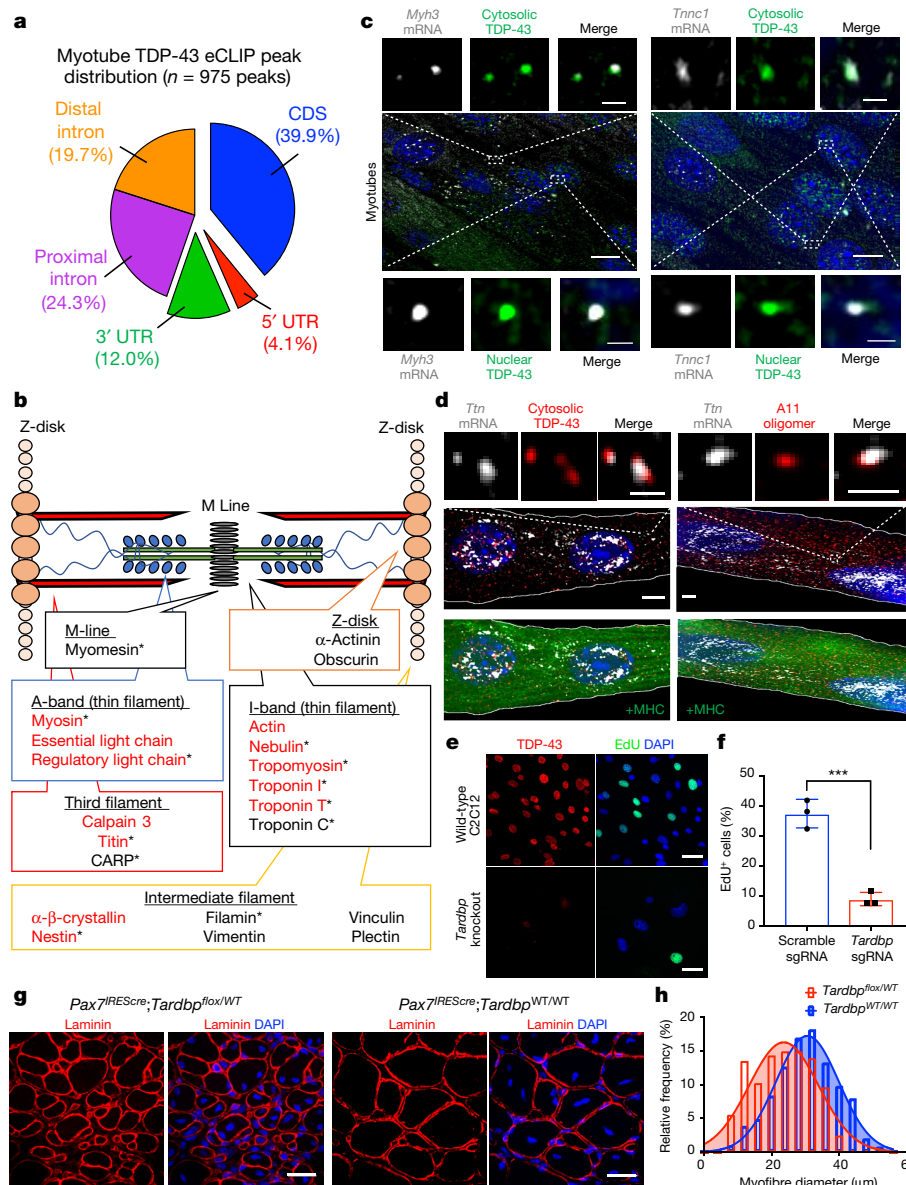


Fig. 3 | TDP-43 binds to select sarcomeric mRNA transcripts during muscle formation. **a**, Distribution of TDP-43 RNA binding identified by eCLIP in C2C12 myotubes. CDS, coding sequence; UTR, untranslated region. **b**, TDP-43 eCLIP identified exonic peaks in select sarcomeric mRNA transcripts in myotubes. All listed genes were found in at least one eCLIP replicate; *gene identified in two replicates; red, gene associated with muscle disease. Sarcomere schematic was adapted from a previous study³⁸. **c**, Single-molecule fluorescence in situ hybridization showed that *Myh3* and *Tnnc1* mRNA co-localized with cytoplasmic and nuclear TDP-43 in C2C12 myotubes. $n = 3$ biologically independent experiments. **d**, Single-molecule fluorescence in situ hybridization showed that *Ttn* mRNA co-localized with both A11 and TDP-43 in the cytoplasm of MHC⁺ C2C12 myotubes. $n = 3$ biologically independent experiments. **c**, **d**, Scale bars, 10 μ m and 0.5 μ m (insets). **e**, Representative images of C2C12 cells edited using CRISPR-Cas9 and *Tardbp* scramble single-guide (sg)RNA

(top) and *Tardbp* knockout sgRNA (bottom), showing TDP-43 expression and EdU incorporation. Scale bar, 50 μ m. Cells were counterstained with DAPI. $n = 3$ biologically independent experiments, each showing similar results. **f**, Quantification of EdU incorporation in *Tardbp* knockout (*Tardbp* sgRNA) and scramble-sgRNA-treated C2C12 cells after seven days in culture. $n = 3$ independent experiments. Unpaired, two-tailed Student's t -test *** $P = 0.0007$. Data are mean \pm s.d. **g**, Representative images of regenerating tibialis anterior muscle at 10 DPI showing a reduction in the myofibre feret diameter in TDP-43-haploinsufficient Pax7^{iresCre}*Tardbp*^{flx/WT} mice. Laminin identifies myofibres and nuclei are counterstained with DAPI. Scale bars, 50 μ m. $n = 3$ mice per condition. **h**, Frequency distribution of myofibre feret diameters in Pax7^{iresCre}*Tardbp*^{flx/WT} mice at 10 DPI compared to Pax7^{iresCre}*Tardbp*^{WT/WT} controls. More than 450 myofibres were quantified from $n = 3$ mice per genotype.

diseases may promote TDP-43 aggregation. Indeed, cytoplasmic TDP-43 aggregates in skeletal-muscle diseases are often seen in myofibres with centrally located nuclei, which is a hallmark of regeneration^{1,28}. Therefore, we tested whether cytoplasmic myo-granules accumulate in newly regenerated myofibres of *Vcp* mutant mice, a model of multisystem proteinopathy and inclusion body myopathy characterized by TDP-43 aggregation²⁹. When uninjured *Vcp* mutant and wild-type mice were treated with 5-ethynyl-2'-deoxyuridine (EdU), which identifies actively regenerating myofibres that contained newly fused nuclei

arising from muscle stem cells, *Vcp* mutant mice possessed more EdU⁺ centrally located myonuclei compared to *Vcp* wild-type mice (Extended Data Fig. 9b, c). Moreover, in the myofibres with EdU⁺ centrally located nuclei, we detected increased cytoplasmic TDP-43 and A11 amyloid oligomer staining, correlating cytoplasmic TDP-43 aggregation with increased muscle regeneration in *Vcp* mutant mice (Fig. 5a, b and Extended Data Fig. 9d, e).

Consistent with the hypothesis that myo-granules may seed the aggregates seen in disease, myo-granules isolated from C2C12

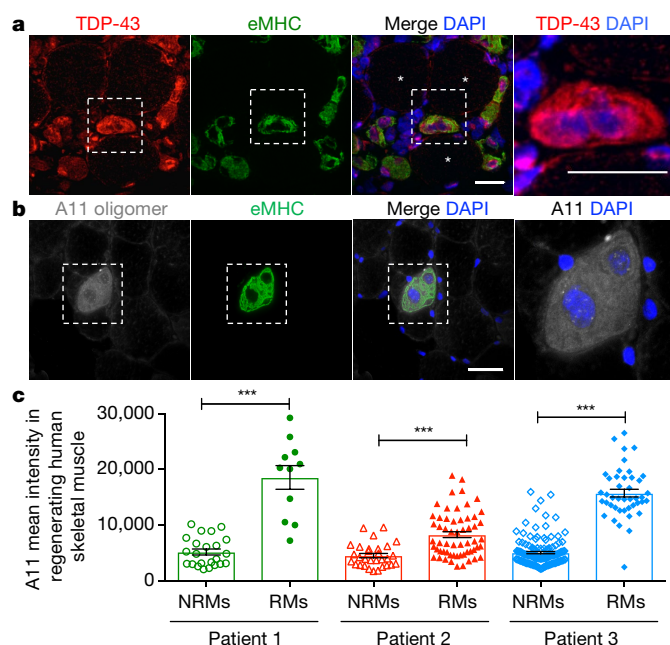


Fig. 4 | Myo-granules form during human muscle regeneration.

a, Representative images of cytoplasmic TDP-43 in regenerating human skeletal muscle from a patient with necrotizing myopathy. Asterisks, uninjured myofibers that lack eMHC and TDP-43 cytosolic signals. $n = 3$ individual patient skeletal muscle biopsies, each showing similar results. Scale bars, 50 μm . **b**, Representative image of A11 immunoreactivity in regenerating human skeletal muscle from a patient with necrotizing myopathy. $n = 3$ individual patient skeletal muscle biopsies, each showing similar results. Scale bar, 100 μm . **c**, Quantification of A11 immunoreactivity in eMHC⁺ regenerating myofibers (RMs) compared to eMHC⁻ non-regenerating myofibers (NRMs) from three patients with necrotizing myopathy. Unpaired, two-tailed Student's *t*-tests were used for each individual comparison: patient 1, NRMs ($n = 23$) versus RMs ($n = 11$), $***P = 2.54 \times 10^{-9}$; patient 2, NRMs ($n = 31$) versus RMs ($n = 59$), $***P = 7.89 \times 10^{-6}$; patient 3, NRMs ($n = 146$) versus RMs ($n = 44$) $***P = 6.17 \times 10^{-49}$. Data are mean \pm s.e.m.

myotubes were capable of transitioning to a thioflavin-T⁺ aggregate (amyloid-like fibres) over time (Fig. 5c, d). Moreover, addition of recombinant TDP-43 to isolated myo-granules increased the amount of thioflavin-T⁺ aggregates that were formed without affecting their initial rate of assembly (Fig. 5c, d and Extended Data Fig. 9f, g). TEM of thioflavin-T⁺ TDP-43 aggregates formed from myo-granules reveals fibrous structures that are morphologically similar to previously reported TDP-43 amyloid fibres³⁰ (Fig. 5e). This suggests that the failure to dissipate myo-granules during normal muscle formation may seed the formation of cytoplasmic TDP-43 aggregates in diseased muscle. Whether the oligomerization of TDP-43 in myo-granules involves its N-terminal oligomerization domain²⁰, the C-terminal prion-like domain that is prone to aggregation and fibre formation^{31,32}, or both, remains to be established.

Discussion

We uncover two important properties of TDP-43 in the formation of skeletal muscle. First, TDP-43 is an essential protein that associates with select sarcomeric mRNAs and localizes to sites of newly forming sarcomeres during skeletal muscle formation. Second, TDP-43 is a component of a higher-order, amyloid-like myo-granule assembled during normal skeletal-muscle formation. Purified myo-granules from cultured myotubes are capable of seeding amyloid-like fibrils in vitro, which suggests a link between the normal biological functions of TDP-43 and pathological TDP-43 aggregates.

We propose a model in which myo-granules that contain TDP-43 are increased in damaged tissues with elevated regeneration, thereby enhancing the possibility of amyloid fibre formation and/or

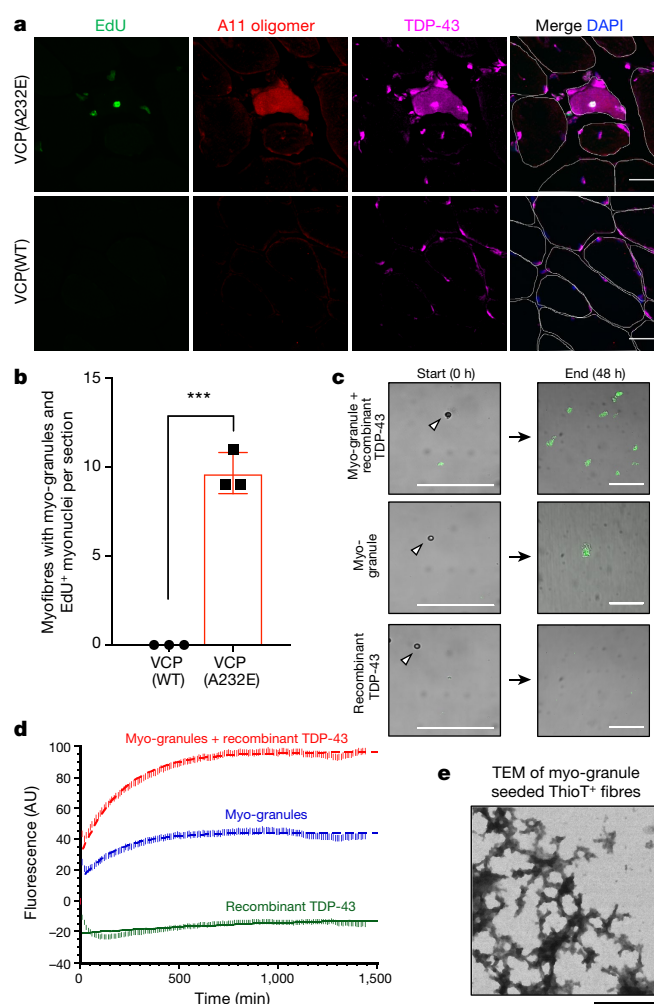


Fig. 5 | Myo-granules are increased in multisystem proteinopathy and are capable of seeding amyloid-like fibres.

a, Tibialis anterior muscles from uninjured VCP(A232E) mice (top) and uninjured VCP wild-type (WT) mice (bottom) analysed for EdU incorporation into centrally located nuclei and immunostained for A11 and TDP-43. Cells were counterstained with DAPI and myofibers are outlined in white. Scale bars, 25 μm . **b**, Quantification of myofibers with EdU⁺ centrally located myonuclei, A11 immunoreactivity and cytoplasmic TDP-43 expression in VCP(A232E) and VCP(WT) mice. Unpaired, two-tailed Student's *t*-test $***P = 1.3 \times 10^{-4}$, $n = 3$ mice, one tibialis anterior cross-section was quantified per mouse. Data are mean \pm s.d. and individual mice are shown. **c**, Representative images of purified myo-granules from C2C12 myotubes incubated with or without recombinant TDP-43 and Thioflavin-T (ThioT) reveals the formation of higher-order thioflavin-T⁺ amyloid-like fibres. Arrowhead points to a bead used to determine correct focal plane at time = 0. Scale bars, 25 μm . $n = 3$ biologically independent experiments. **d**, Plot of kinetics of fibre aggregation determined by thioflavin-T incorporation measured at 10-min intervals. Rates were derived by fitting time points to a single exponential rate (equation (1); see Methods). Myo-granule + recombinant TDP-43, $R^2 = 0.96$, $k_{\text{observed}} = 47 \pm 1.6 \times 10^{-4} \text{ min}^{-1}$ (mean \pm s.d.) myo-granule, $R^2 = 0.92$, $k_{\text{observed}} = 56 \pm 2.9 \times 10^{-4} \text{ min}^{-1}$; recombinant TDP-43, $R^2 = 0.47$, $k_{\text{observed}} = 8.5 \pm 4.9 \times 10^{-4} \text{ min}^{-1}$. $n = 3$ biologically independent experiments, background-corrected arbitrary units (AU). **e**, Representative TEM images of thioflavin-T⁺ fibres formed from isolated myo-granules. $n = 3$ biologically independent experiments. Scale bar, 1 μm .

aggregation of TDP-43 in disease (Extended Data Fig. 10). Because the triggering event in this model is elevated muscle regeneration, it explains why TDP-43 aggregates occur in genetically diverse diseases, including inclusion body myopathy²⁸, which can be caused by mutations in the ubiquitin segregase VCP²⁹; oculopharyngeal muscular dystrophy, caused by Ala expansions in PABPN1¹; and distal

myopathy with rimmed vacuoles, caused by mutations in the UDP-N-acetylglucosamine 2-epimerase gene (*GNE*)³³. Moreover, the seeding of TDP-43 aggregates by TDP-43 oligomers may also occur in neurons as reversible cytoplasmic TDP-43 accumulation occurs in models of acute neuronal injury in vivo (for example, axotomy or traumatic brain injury)^{34,35}. TDP-43 aggregates are also frequently observed on autopsy in neurologically normal elderly individuals³⁶. The age-dependent accumulation of TDP-43 aggregates may be caused by a failure to clear TDP-43, or other amyloid-like assemblies that have formed during tissue repair. Over a lifetime, failures in proteostatic control mechanisms—including autophagy or endocytosis³⁷—could increase the likelihood that functional, amyloid-like assemblies transition into pathological aggregates.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0665-2>.

Received: 5 January 2018; Accepted: 3 October 2018;

Published online 31 October 2018.

- Küsters, B. et al. TDP-43 accumulation is common in myopathies with rimmed vacuoles. *Acta Neuropathol.* **117**, 209–211 (2009).
- Weihl, C. C. et al. TDP-43 accumulation in inclusion body myopathy muscle suggests a common pathogenic mechanism with frontotemporal dementia. *J. Neurol. Neurosurg. Psychiatry* **79**, 1186–1189 (2008).
- Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
- Renton, A. E., Chiò, A. & Traynor, B. J. State of play in amyotrophic lateral sclerosis genetics. *Nat. Neurosci.* **17**, 17–23 (2014).
- Kraemer, B. C. et al. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathol.* **119**, 409–419 (2010).
- Schmid, B. et al. Loss of ALS-associated TDP-43 in zebrafish causes muscle degeneration, vascular dysfunction, and reduced motor neuron axon outgrowth. *Proc. Natl Acad. Sci. USA* **110**, 4986–4991 (2013).
- Diaper, D. C. et al. *Drosophila* TDP-43 dysfunction in glia and muscle cells cause cytological and behavioural phenotypes that characterize ALS and FTL. *Hum. Mol. Genet.* **22**, 3883–3893 (2013).
- Llamusi, B. et al. Muscleblind, BSF and TBPH are mislocalized in the muscle sarcomere of a *Drosophila* myotonic dystrophy model. *Dis. Model. Mech.* **6**, 184–196 (2013).
- Rodriguez-Ortiz, C. J. et al. Neuronal-specific overexpression of a mutant valosin-containing protein associated with IBMPFD promotes aberrant ubiquitin and TDP-43 accumulation and cognitive dysfunction in transgenic mice. *Am. J. Pathol.* **183**, 504–515 (2013).
- Caldwell, C. J., Matthey, D. L. & Weller, R. O. Role of the basement membrane in the regeneration of skeletal muscle. *Neuropathol. Appl. Neurobiol.* **16**, 225–238 (1990).
- Hardy, D. et al. Comparative study of injury models for studying muscle regeneration in mice. *PLoS ONE* **11**, e0147198 (2016).
- Webster, C., Silberstein, L., Hays, A. P. & Blau, H. M. Fast muscle fibers are preferentially affected in Duchenne muscular dystrophy. *Cell* **52**, 503–513 (1988).
- Johnson, B. S. et al. TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J. Biol. Chem.* **284**, 20329–20339 (2009).
- Sangwan, S. et al. Atomic structure of a toxic, oligomeric segment of SOD1 linked to amyotrophic lateral sclerosis (ALS). *Proc. Natl Acad. Sci. USA* **114**, 8770–8775 (2017).
- Kayed, R. et al. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486–489 (2003).
- Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
- Ayala, Y. M. et al. TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* **30**, 277–288 (2011).
- Polymenidou, M. et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* **14**, 459–468 (2011).
- Tollervey, J. R. et al. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **14**, 452–458 (2011).
- Afroz, T. et al. Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. *Nat. Commun.* **8**, 45 (2017).
- Alami, N. H. et al. Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. *Neuron* **81**, 536–543 (2014).
- Freibaum, B. D., Chitta, R. K., High, A. A. & Taylor, J. P. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *J. Proteome Res.* **9**, 1104–1120 (2010).
- El Fatimy, R. et al. Tracking the fragile X mental retardation protein in a highly ordered neuronal ribonucleoprotein population: a link between stalled polyribosomes and RNA granules. *PLoS Genet.* **12**, e1006192 (2016).
- Taylor, J. P. Multisystem proteinopathy: intersecting genetics in muscle, bone, and brain degeneration. *Neurology* **85**, 658–660 (2015).
- Kim, H. J. et al. Mutations in prion-like domains in *hnrnpA2B1* and *hnrnpA1* cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
- Chiang, P.-M. et al. Deletion of *TDP-43* down-regulates *Tbc1d1*, a gene linked to obesity, and alters body fat metabolism. *Proc. Natl Acad. Sci. USA* **107**, 16320–16324 (2010).
- Murphy, M. M., Lawson, J. A., Mathew, S. J., Hutcheson, D. A. & Kardon, G. Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *Development* **138**, 3625–3637 (2011).
- Salajegheh, M. et al. Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis. *Muscle Nerve* **40**, 19–31 (2009).
- Custer, S. K., Neumann, M., Lu, H., Wright, A. C. & Taylor, J. P. Transgenic mice expressing mutant forms VCP/p97 recapitulate the full spectrum of IBMPFD including degeneration in muscle, brain and bone. *Hum. Mol. Genet.* **19**, 1741–1755 (2010).
- Mompeán, M. et al. Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J. Phys. Chem. Lett.* **6**, 2608–2615 (2015).
- Chen, A. K.-H. et al. Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc.* **132**, 1186–1187 (2010).
- Igaz, L. M. et al. Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. *J. Biol. Chem.* **284**, 8516–8524 (2009).
- Nishino, I. et al. Distal myopathy with rimmed vacuoles is allelic to hereditary inclusion body myopathy. *Neurology* **59**, 1689–1693 (2002).
- Wiesner, D. et al. Reversible induction of TDP-43 granules in cortical neurons after traumatic injury. *Exp. Neurol.* **299**, 15–25 (2018).
- Moisse, K. et al. Divergent patterns of cytosolic TDP-43 and neuronal progranulin expression following axotomy: implications for TDP-43 in the physiological response to neuronal injury. *Brain Res.* **1249**, 202–211 (2009).
- Wilson, R. S. et al. TDP-43 pathology, cognitive decline, and dementia in old age. *JAMA Neurol.* **70**, 1418–1424 (2013).
- Liu, G. et al. Endocytosis regulates TDP-43 toxicity and turnover. *Nat. Commun.* **8**, 2092 (2017).
- Laing, N. G. & Nowak, K. J. When contractile proteins go bad: the sarcomere and skeletal muscle disease. *BioEssays* **27**, 809–822 (2005).

Acknowledgements We thank J. Dragavon, J. Wei Tay, J. Orth and G. Morgan for help with microscopy; C. Glabe for A11 antibodies, P. Wong for *Tardbp*^{flax} mice, T. Elston for help with fluorescence-activated cell sorting, T. Lee for help with mass spectrometry; and M. Wicklund, S. Ringel and S. Reed for work with patient samples. The research was supported by NIH-T32GM008497 (J.R.W.), E.D.N., T.O.V. and E.L.), NIH-F30NS093682 (J.R.W.), NIH-F30AR068881 (T.O.V.), NIH-GM045443 (R.P.), the Howard Hughes Medical Institute (R.P., D.S.E. and J.P.T.), NIH-R35GM119575 (A.M.J.), Paul O'Hara II Seed Grant from ACS-IRG Grant Program (A.M.J.), University of Colorado Cancer Center Genomics Core (supported by NIH-P30CA46934), NIH-AR049446 and NIH-AR070360 (B.B.O.), Glenn Foundation for Biomedical Research (B.B.O.), Beverly Sears Grant (J.R.W.), NSF MCB 1616265 and NIH NIA AG054022 (D.S.E.) and a Butcher Innovation Award NSF IGERT 1144807 (J.R.W. and T.O.V.).

Reviewer information Nature thanks A. D. Gitler, E. Olson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.O.V., J.R.W., B.B.O. and R.P. conceived and designed the research, wrote the manuscript and all authors edited drafts. T.O.V., J.R.W., E.L., N.D.B. and O.N.W. performed and analysed mouse regeneration and myotube formation experiments. J.R.W. and E.L. isolated myo-granules. T.E.E., J.R.W. and T.O.V. analysed HaloTag-TDP-43. M.P.H., J.R.W. and T.O.V. performed X-ray diffraction and TEM. E.D.N. and J.R.W. performed eCLIP analysis. T.O.V. and K.A.B. analysed human biopsies. T.O.V. performed VCP experiments. J.R.W. and B.R. performed thioflavin-T assays. E.G., J.S., T.E.L., D.S.E., J.P.T. and A.M.J. provided scientific insights and materials.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0665-2>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0665-2>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to B.B.O. or R.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mice. Mice were bred and housed according to National Institutes of Health (NIH) guidelines for the ethical treatment of animals in a pathogen-free facility at the University of Colorado at Boulder (wild-type, *Pax7^{IREScree}*, *Tardbp^{fllox/fllox}* and VCP(A232E) lines). The University of Colorado Institutional Animal Care and Use Committee (IACUC) approved all animal protocols and procedures and studies complied with all ethical regulations. Wild-type mice were C57BL/6 (Jackson Laboratories) and VCP(WT)²⁹, and *Tardbp^{fllox/fllox}* mice²⁶ were previously described. Crossing mice into *Pax7^{IREScree}* mice²⁷ generated conditional *Tardbp^{fllox/WT}* mice. Cells and tibialis anterior muscles were isolated from 3–6-month-old male and female wild-type and *Pax7^{IREScree}* *Tardbp^{fllox/WT}* mice. Tibialis anterior or gastrocnemius muscles were isolated from nine-month-old male VCP(A232E) mice. Control mice were randomly assigned and were age- and sex-matched to the mice and crosses described above. Sample sizes were set at $n = 3$ unless otherwise noted. No statistical methods were used to predetermine sample size.

Mouse injuries and tamoxifen injections. Mice at 3–6 months old were anaesthetized with isoflurane and the left tibialis anterior muscle was injected with 50 μ l of 1.2% BaCl₂ and then the injured and contralateral tibialis anterior muscles were collected at the indicated time points. Tamoxifen (Sigma-Aldrich), resuspended in sterile corn oil (Sigma-Aldrich), was administered by intraperitoneal injection to 3–6-month-old mice at a volume of 0.075 mg of tamoxifen per gram of mouse weight. Muscle injuries were made blinded to genotype.

Human muscle biopsy tissue. Under an IRB-approved protocol at Johns Hopkins University and complying with all ethical regulations, a clinical muscle biopsy database was searched for patients who had been clinically diagnosed with rhabdomyolysis and/or pathologically diagnosed with necrotizing myopathy with evidence of myofibre regeneration. Muscle biopsy specimens used in this study were left over from diagnostic biopsies and the IRB approved that patient consent was not necessary. Patient muscle tissue leftover from the diagnostic biopsy was stored frozen at -80°C for less than two years, and samples were cryo-sectioned for immunohistochemical analysis.

Immunofluorescence staining of tissue sections. Tibialis anterior or gastrocnemius muscles were dissected, fixed on ice for 2 h with 4% paraformaldehyde, and then transferred to phosphate-buffered saline (PBS) with 30% sucrose at 4°C overnight. Muscle was mounted in OCT (Tissue-Tek) and cryo-sectioning was performed on a Leica cryostat to generate 10- μ m thick sections. Tissues and sections were stored at -80°C until staining. Tissue sections were post-fixed in 4% paraformaldehyde for 10 min at room temperature and washed three times for 5 min in PBS. Immunostaining with anti-PAX7, anti-laminin, anti-eMHC, anti-TDP-43 and A11 antibodies required heat-induced epitope retrieval, for which post-fixed slides were placed in citrate buffer, pH 6.0, and subjected to 6 min of high pressure-cooking in a Cuisinart model CPC-600 pressure cooker. For immunostaining, tissue sections were permeabilized with 0.25% Triton X-100 (Sigma-Aldrich) in PBS containing 2% bovine serum albumin (BSA) (Sigma-Aldrich) for 60 min at room temperature. Incubation with primary antibody occurred at 4°C overnight followed by incubation with the secondary antibody at room temperature for 1 h. Primary antibodies included mouse anti-PAX7 (Developmental Studies Hybridoma Bank) at 1:750, rabbit anti-laminin (Sigma-Aldrich) at 1:200, rabbit anti-TDP-43 (ProteinTech) at 1:200, mouse anti-TDP-43 (Abcam) at 1:200, rabbit A11 (Sigma-Aldrich) at 1:200, mouse anti-VCP (ThermoFisher Scientific) at 1:400 and mouse anti-eMHC (Developmental Studies Hybridoma Bank) at 1:5. Alexa secondary antibodies (Molecular Probes) were used at a dilution of 1:1,000. For analysis that included EdU detection, EdU staining was completed before antibody staining using the Click-iT EdU Alexa Fluor 488 detection kit (Molecular Probes) following the manufacturer's protocols. Sections were incubated with 1 μ g ml⁻¹ DAPI for 10 min at room temperature then mounted in Mowiol supplemented with DABCO (Sigma-Aldrich) or ProLong Gold (Thermo) as an anti-fade agent.

Isolation of primary muscle stem cells. Gastrocnemius, extensor digitorum longus, tibialis anterior and all other lower hindlimb muscles were dissected from wild-type mice. The muscle groups from both hindlimbs were separated and digested in 3.6 ml of F12-C (Gibco) with penicillin and streptomycin (Gibco) and 400 μ l 10 \times collagenase (Worthington) for 90 min at 37°C on a slow rotisserie. In a biosafety cabinet, muscles allowed to settle for 5 min, undisturbed. Then, as much of the liquid as possible was removed without disturbing the muscle groups. F12-C with penicillin and streptomycin and 15% horse serum was added and the muscles were rocked for 1 min. Then, 3 ml of growth medium was added to each tube (F12-C with penicillin and streptomycin, 15% horse serum (Gibco), 20% fetal bovine serum (Sigma-Aldrich), 1% chick embryo extract (Antibody Production Services). The digest was poured onto a 10-cm tissue-culture plate (Corning) with Matrigel in 10 ml of growth medium. Growth medium was added as necessary to keep the muscle chunks submerged. Muscles chunks were incubated in growth medium with FGF-2 (50 nM working concentration) for 72 h at 37°C in 6% O₂ and 5% CO₂. After 72 h, muscle stem cells had migrated out onto the Matrigel and

the muscle chunks and medium were removed. The plate containing attached muscle stem cells was rinsed with sterile PBS and 10 ml warm growth medium was added supplemented with 50 nM FGF-2. Colonies of myoblasts formed by four days of culture and were expanded by passaging with 0.25% trypsin-EDTA (Sigma-Aldrich).

Cell culture. *Primary muscle stem cells.* After initial isolation, primary myoblasts were maintained on Matrigel-coated tissue-culture plastic plates or gelatin-coated coverslips at 37°C at 6% O₂ and 5% CO₂ in growth medium as described above. Medium was changed only during cell passaging. To promote myoblast fusion, cells at 75% confluency were washed three times with PBS and the medium was switched to DMEM (Gibco) with 5% horse serum (Gibco), 1% penicillin and streptomycin and 1% insulin–transferrin–selenium (Gibco). To induce stress-granule formation, primary myotubes were stressed with 0.5 mM sodium arsenite for 1 h at 37°C .

C2C12 cells. Immortalized mouse myoblasts (American Type Culture Collection) were maintained on uncoated standard tissue-culture plastic or gelatin-coated coverslips at 37°C with 5% CO₂ in DMEM with 20% fetal bovine serum and 1% penicillin and streptomycin. To promote myoblast fusion when the C2C12 cells reached confluence, they were switched to 5% horse serum, 1% penicillin and streptavidin and 1% insulin–transferrin–selenium in DMEM. To induce stress-granule formation, C2C12 myotubes were stressed with 0.5 mM sodium arsenite for 1 h at 37°C .

U2-OS cells. Human osteosarcoma cells were maintained in DMEM, high glucose, GlutaMAX with 10% fetal bovine serum, 1% penicillin–streptomycin and 1 mM sodium pyruvate at 37°C and 5% CO₂.

Yeast. For the experiments presented in Fig. 1, BY4741 yeast was transformed with a single plasmid expressing Pub1Q/N-GFP (pRP1689) (laboratory of R.P.) and grown at 30°C in minimal medium with 2% glucose as a carbon source and with leucine dropout to maintain the plasmid. For experiments presented in Supplementary Fig. 3, SUP35 [PSI⁺] (5V-H19A) and SUP35 [psi⁻] (yAV831) strains were grown in minimal medium supplemented with a complete set of amino acids and 2% dextrose at 30°C .

Immunofluorescence staining of cells and proximity ligation assay. Primary and immortalized cells were washed with PBS in a laminar flow hood and fixed with 4% paraformaldehyde for 10 min at room temperature in a chemical hood. Cells were permeabilized with 0.25% Triton X-100 in PBS containing 2% BSA (Sigma-Aldrich) for 1 h at room temperature. Incubation with primary antibody occurred at 4°C overnight followed by incubation with the secondary antibody at room temperature for 1 h. Primary antibodies included mouse anti-PAX7 (Developmental Studies Hybridoma Bank) at 1:750, rabbit anti-TDP-43 (ProteinTech) at 1:200, mouse anti-TDP-43 (Abcam) at 1:200, rabbit A11 (Sigma-Aldrich) at 1:200 and mouse anti-MHC (MF-20, Developmental Studies Hybridoma Bank) at 1:1. Alexa secondary antibodies (Molecular Probes) were used at a dilution of 1:1,000. All antibodies were diluted in 0.125% Triton X-100 in PBS containing 2% BSA. For analysis that included EdU detection, EdU staining was completed before antibody staining using the Click-iT EdU Alexa Fluor 488 detection kit (Molecular Probes) following the manufacturer's protocol. Cells were incubated with 6.6 mM phalloidin (Thermo Scientific) for 20 min and/or 1 μ g ml⁻¹ DAPI for 10 min at room temperature then mounted in Mowiol supplemented with DABCO (Sigma-Aldrich) as an anti-fade agent.

For the proximity ligation assay, samples were incubated with indicated antibodies at the concentrations listed above. Secondary antibody incubation and Duolink proximity ligation assays were performed according to the manufacturer's protocol (Sigma-Aldrich).

Subcellular fractionation. Nuclear/cytosolic fractionation was performed to determine localization of soluble TDP-43 in C2C12 myoblasts and differentiating myotubes. In brief, myoblasts or differentiating myotubes (day 4) were trypsinized, washed with PBS and pelleted by centrifugation at 1,000g for 5 min. Cells were subsequently washed in PBS and divided into a whole-cell lysate fraction (1/3 total) or a cytosolic/nuclear fraction (2/3 total). Both cellular fractions were pelleted by centrifugation at 1,000g for 5 min. The whole-cell lysate fraction was resuspended into RIPA buffer (50 mM Tris pH 7.5, 1% NP-40, 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl and protease inhibitors (Roche)) and placed on ice. The cytosolic/nuclear fraction was resuspended in a hypotonic lysis buffer (10 mM Tris HCL 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.5% NP40 and protease inhibitors (Roche)) and placed on ice for 4 min. Nuclei were then pelleted by centrifugation at 500g for 5 min. The supernatant (cytosolic fraction) was removed. The pellet (nuclear fraction) was then resuspended in nuclear lysis buffer (50 mM Tris HCL 7.4, 120 mM NaCl, 1% SDS, 1 mM EDTA, 50 mM DTT and protease inhibitors (Roche)). Nuclei were lysed with five passages through an 18G needle. Cellular debris was cleared from collected fractions with centrifugation at 1,000g for 5 min. Equal volumes (20 μ l) of fractions were then resolved on a 4–12% Bis-Tris SDS–PAGE gel and transferred to a nitrocellulose membrane (Bio-Rad). Western blotting was performed according to standard procedures.

Single molecule imaging of endogenous HaloTag–TDP-43. A tetracycline-inducible HaloTag (Promega) TDP-43 fusion protein was knocked into the *Rosa26* safe-harbour locus using CRISPR–Cas9³⁹. Knockin cells were selected using puromycin and proper genomic integration was confirmed by PCR and western blotting. For live-cell single-molecule imaging studies, puromycin-resistant myoblasts or differentiating myotubes were grown on collagen-treated, 35-mm imaging dishes (MatTek). HaloTag–TDP-43 was induced for 48 h using doxycycline ($1 \mu\text{g ml}^{-1}$). HaloTag–TDP-43 molecules were labelled with 50 pM JF646 dye (gift from L. Lavis) for 15 min in culture medium⁴⁰. After the pulse, cells were washed three times with medium and incubated with vibrant violet (1:400) in medium to visualize myonuclei for at least 1 h before image acquisition. All single-molecule live imaging was performed under HILO conditions (highly inclined and laminated optical sheet) on a Nikon N-STORM microscope equipped with TIRF illuminator, an environmental chamber, two iXon Ultra 897 EMCCD cameras (Andor), a $100\times$ oil-immersion objective (Nikon, NA 1.49), two filter wheels, appropriate filter sets, and 405 nm (20 mW), 488 nm (50 mW), 561 nm (50 mW), and 647 nm (125 mW) laser lines. Differentiating myotubes were identified by visualizing fused myonuclei with a 405 nm laser line (1% laser power). To image HaloTag–TDP-43, cells were imaged continuously with 647 nm (40% laser power) for 15 s at an effective frame rate of 100 frames per s. Single-particle tracks were generated using MATLAB.

Biochemical characterization of TDP-43 during myogenesis. For RIPA/urea solubility assays, C2C12 myoblasts and myotubes were lysed with RIPA buffer (50 mM Tris pH 7.5, 1% NP-40, 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl). Protein concentrations were determined using BCA assay (Thermo Scientific) according to standard procedures. Lysates were centrifuged at 18,000g for 20 min at 4°C. The supernatant represented the RIPA-soluble fraction while the pellet was solubilized in 7 M urea in TBE and represents the urea-soluble fraction. Western blotting was performed following resolution of protein lysates on SDS–PAGE.

Semi-denaturing detergent–agarose gel electrophoresis (SDD–AGE) was conducted as previously described⁴¹. In brief, C2C12 myoblasts and myotubes were lysed with RIPA buffer, protein concentrations were standardized using BCA assay, diluted to $1\times$ in loading buffer ($2\times$ TAE, 20% glycerol, 8% SDS and bromophenol blue) and separated across a 1.5% agarose gel containing 0.1% SDS. Gels were transferred by capillary transfer overnight to nitrocellulose in TBS. Standard western blotting procedures were used.

For fractionation of TDP-43 oligomers across sucrose gradients (10–35%), fractions were collected and equal volumes were loaded for SDD–AGE analysis. Immunoprecipitation followed by scanning electron microscopy of the TDP-43 SDS-resistant fraction was performed as described below.

Dot blots of C2C12 protein lysates or whole-muscle lysates were conducted according to standard procedures⁴². Both C2C12 cells and whole muscle were lysed in RIPA buffer, protein concentrations were normalized using BCA and were spotted onto nitrocellulose membranes.

Isolation of myo-granules. A protocol for isolating myo-granules from myotubes was modified from existing protocols for isolating heavy ribonucleoprotein complexes⁴³. In brief, myotubes or whole tibialis anterior muscles were lysed under non-denaturing conditions using CHAPS lysis buffer (10 mM Tris-HCl pH 7.5, 1 mM MgCl_2 , 1 mM EGTA, 0.5% CHAPS, 10% glycerol, 1 mM PMSF and 1 mM DTT) or RIPA buffer and spun to remove heavy cellular debris (250g for 5 min). Successive centrifugation was used to enrich for heavy complexes (18,000g for 20 min). The pellet was resuspended into immunoprecipitation buffer (10 mM Tris HCl 7.5, 25 mM NaCl and 0.005% NP40) to create the ‘myo-granule-enriched fraction’. The enriched fraction was precleared for 30 min with immunoprecipitation buffer-equilibrated Dynabeads and then incubated overnight with either antibodies against TDP-43 (Proteintech) or the A11 antibody (laboratory of C. Glabe). Myo-granules were immunopurified on equilibrated Dynabeads, washed in immunoprecipitation buffer, and eluted using Pierce Gentle Ag/Ab Buffer (Thermo Scientific) as previously described⁴². Buffer was exchanged using a 10K MW spin column (Millipore Amicon).

RNA extraction and oligo-dT northern blot analysis of myo-granules. RNA was isolated from myo-granules bound to Dynabeads by Trizol extraction followed by ethanol precipitation. RNA was run on a 1.25% formaldehyde agarose gel, transferred to nitrocellulose membrane and hybridized with a αP^{32} -labelled oligo-dT probe at room temperature overnight. Membranes were exposed on a phosphorimager screen either for 1 h (low exposure) or overnight (high exposure) and imaged on a Typhoon FLA 9500 phosphorimager.

TEM. TEM sample preparation and image acquisition was performed as previously described unless otherwise specified⁴⁴. For experiments in which immunofluorescence on TEM grids was performed, Carbon type B 300 mesh Copper TEM grids (Ted Pella) were poly-lysine-treated (Sigma-Aldrich) for 30 min, washed three times in PBS, and immunopurified myo-granules (diluted 1:50) were allowed to adhere to the grid for 1 h at room temperature. TEM grids with myo-granules were

blocked in 3% BSA for 1 h at room temperature. Primary antibody incubation was performed at a dilution of 1:100 in 3% BSA for 1 h at room temperature. Grids were then washed three times in PBS and incubated with secondary antibodies at 1:250 dilution in 3% BSA. Secondary antibody-only controls were performed at the same concentration without addition of primary antibodies. Grids were washed three times with PBS and placed onto microscopy slides. Images were acquired using a DeltaVision Elite microscope with a $100\times$ objective. Grids were stained with uranyl acetate and immunopositive myo-granules were examined by TEM. **Myo-granule electron diffraction.** Lyophilized myo-granules and SOD1 segment oligomers (prepared as previously described¹⁴) were mounted for diffraction by dipping a nylon loop in glycerol and sticking some of the lyophilized sample onto the glycerol⁴⁵. Samples were carefully aligned to avoid the nylon loop entering the X-ray beam when diffraction images were taken. All samples were shot at the Advanced Photon Source (Argonne National Laboratory) beamline 24-E with a $50\text{-}\mu\text{m}$ aperture. Samples were rotated 5 degrees over a 4-s exposure at 295 K and images were analysed with ADXV.

TDP-43 eCLIP sequencing. C2C12 myoblasts were seeded at 6×10^6 cells per 15-cm plate, grown for 24 h at 37°C, 5% CO_2 and either collected (undifferentiated myoblasts) or differentiated in differentiation medium for seven days. TDP-43 eCLIP was performed according to established protocols¹⁶.

In brief, TDP-43–RNA interactions were stabilized with ultraviolet-light crosslinking (254 nm, 150 mJ cm^{-2}). Cell pellets were collected and snap-frozen in liquid nitrogen. Cells were thawed, lysed in eCLIP lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate and $1\times$ protease inhibitor) and sonicated (Bioruptor). Lysate was RNase I-treated (Ambion, 1:25) to fragment RNA. Protein–RNA complexes were immunoprecipitated using the indicated antibody. One size-matched input library was generated per biological replicate using an identical procedure without immunoprecipitation. Stringent washes were performed as described, RNA was dephosphorylated (FastAP, Fermentas), T4 PNK (NEB), and a 3′-end RNA adaptor was ligated with T4 RNA ligase (NEB). Protein–RNA complexes were resolved on an SDS–PAGE gel, transferred to nitrocellulose membranes and RNA was extracted from membrane. After RNA precipitation, RNA was reverse-transcribed using SuperScript IV (Thermo Fisher Scientific), free primers were removed, and a 3′ DNA adaptor was ligated onto cDNA products with T4 RNA ligase (NEB). Libraries were PCR-amplified and dual-indexed (Illumina TruSeq HT). Pair-end sequencing was performed on Illumina NextSeq sequencer.

Bioinformatics and statistical analysis. Read processing and cluster analysis for TDP-43 eCLIP was performed as previously described¹⁶. Read processing and cluster analysis for TDP-43 eCLIP was performed as previously described. In brief, 3′ barcodes and adaptor sequences were removed using standard eCLIP scripts. Reads were trimmed, filtered for repetitive elements and aligned to the mm9 reference sequence using STAR. PCR duplicate reads were removed based on the read start positions and random sequence. Bigwig files for genome browser display were generated based on the location of the second of the paired-end reads. Peaks were identified using the encode_branch version of CLIPPER using the parameter ‘-s mm9’. Peaks were normalized against size-matched input by calculating fold enrichment of reads in immunoprecipitated samples versus input, and were deemed significant if the number of reads in the immunoprecipitated sample was greater than in the input sample, with a Bonferroni-corrected Fisher exact *P* value less than 10^{-8} .

Microscopy and image analyses. Images were captured on a Nikon inverted spinning disk confocal microscope or a DeltaVision Elite microscope. Objectives used on the Nikon were: $10\times/0.45$ NA Plan Apo, $20\times/0.75$ NA Plan Apo and $40\times/0.95$ NA Plan Apo. Confocal stacks were projected as maximum intensity images for each channel and merged into a single image. Brightness and contrast were adjusted for the entire image as necessary. Both muscle stem cell numbers and average fibre diameter were counted manually using Fiji ImageJ. Objectives used on the DeltaVision Elite microscope were $100\times$ using a PCO Edge sCMOS camera. At least three images were taken for each experiment comprising 8–10 *z* sections each. Images were processed using Fiji ImageJ. For super-resolution imaging, microscopy was performed using a Leica TCS SP8 White Light Laser with $63\times/1.4$ NA oil objective coupled to HyVolution (SVI Huygens-based deconvolution) and special Leica Hybrid Detectors. Image quantification was performed using Imaris imaging software.

Sequential immunofluorescence and single-molecule fluorescence in situ hybridization. Sequential immunofluorescence and single-molecule fluorescence in situ hybridization on fixed myotubes was performed. In brief, C2C12 myotubes were differentiated for seven days in differentiation medium, fixed in 4% paraformaldehyde (4%) for 10 min and washed in PBS. The following antibodies were used for immunofluorescence: rabbit anti-TDP-43 (Proteintech, 1:400), rabbit anti-A11 oligomer (Thermo Fisher Scientific, 1:400), goat anti-rabbit Alexa 647 (Abcam, 1:1,000), goat anti-mouse IgG1 Alexa 488 (Thermo Fisher Scientific, 1:1,000). All immunofluorescence experiments were performed sequentially except for staining

with mouse anti-myosin heavy chain, F59 (DSHB) which was diluted (1:10) in hybridization buffer. Custom Stellaris FISH probes were designed against mouse *Ttn*, *Myh3*, *Tnnc1* and probes were labelled with Quasar 570 Dye using Stellaris RNA FISH Probe Designer (Biosearch Technologies).

Mass spectrometry. Mass spectrometry (MS) was performed as previously described⁴³. In brief, samples were immunoprecipitated on Dynabeads as described above. Samples were washed with 0.1 M ammonium bicarbonate, and resuspended in 100 μ l of 0.1 M ammonium bicarbonate, 0.2% sodium deoxycholate and 6 M guanidine HCL. Samples were reduced and alkylated with 5 mM TCEP, 40 mM chloroacetamide at 65 °C for 20 min in darkness. Samples were trypsinized with 0.5 μ g of trypsin at 37 °C for overnight. The proteolysis reaction was quenched by acidification using formic acid. Deoxycholic acid was removed by phase-transfer using ethyl acetate. Tryptic peptides were desalted using in-house stop-and-go extraction (STAGE) tips, speed-vac to dryness and samples were stored at –80 °C.

Samples were resolved by ultra-performance liquid chromatography in the direct injection mode using a Waters nanoACQUITY system. Samples were resuspended in 12 μ l of buffer A (0.1% formic acid/water), of which 5 μ l (42% of total) was loaded onto a Symmetry C18 nanoACQUITY trap column (130 Å, 5 μ m, 180 μ m \times 20 mm) with 15 μ l min^{–1} of 99.5% buffer A and 0.5% buffer B (0.1% formic acid/acetonitrile) for 3 min. Samples were then eluted and resolved on a BEH130 C18 analytical column (130 Å, 1.7 μ m, 75 μ m \times 250 mm) using a gradient with 3–8% buffer B between 0 and 3 min, 8–28% buffer B between 2 and 185 min, and 28–60% buffer B between 185 and 190 min (0.3 μ l min^{–1}). MS/MS was performed using an LTQ Orbitrap Velos, scanning mass spectrometry between 400 and 1,800 *m/z* (1 \times 10⁶ ions, 60,000 resolution) in Fourier Transform, and selecting the 20 most intense MH₂²⁺ and MH₃³⁺ ions for MS/MS in linear trap quadrupole with 180 s dynamic exclusion, 10 p.p.m. exclusion width, repeat count = 1. Maximal injection time was 500 ms for FT precursor scans with one microscan, and 250 ms for LTQ-MS/MS with one microscan and automated gain control 1 \times 10⁴. The normalized collision energy was 35%, with activation *Q* = 0.25 for 10 ms.

Raw data from mass spectrometry were processed using MaxQuant/Andromeda (version 1.5.0.12) and searched against the Uniprot mouse database (downloaded in October 2015, 46,471 entries) with common contaminant entries. The search used trypsin specificity with maximum two missed cleavages, included carbamidomethylation on Cys as a fixed modification, and N-terminal acetylation and oxidation on Met as variable modifications. Andromeda used 7 p.p.m. maximum mass deviation for the precursor ion, and 0.5 Da as MS/MS tolerance, searching eight top MS/MS peaks per 100 Da. False discovery rates were set to 0.01 for both protein and peptide identifications, with a minimum peptide length of seven amino acids, and two minimum total peptides.

Tardbp CRISPR–Cas9 knockout and EdU incorporation. CRISPR–Cas9 knockout was performed in C2C12 myoblasts. sgRNAs against TDP-43 (5′-GTGTATGAGAGGAGTCCGAC-3′) were designed using CRISPR Design version 1 (<http://crispr.mit.edu/>) and cloned into pSpCas9(BB)-2A-Puro (PX459). T7 endonuclease assays was used to confirm correct targeting to the *Tardbp* locus. C2C12 myoblasts were transfected with JetPrime using standard protocols. Myoblasts were selected with puromycin (1 μ g ml^{–1}) for one week. C2C12 myoblasts were incubated with 10 μ M EdU (Life Technologies) for 3 h. Cells were washed, fixed and stained using the methods described above.

Recombinant TDP-43 purification. Full-length human TDP-43 was subcloned into pE-SUMO (LifeSensors). His6–SUMO N-terminally tagged TDP-43 was transformed in BL21(DE3)RIL *Escherichia coli*, which were grown up from an overnight culture in LB containing ampicillin at 37 °C until an optical density at

600 nm (OD₆₀₀) of 0.3 was reached. At this time, the culture was shifted to 15 °C and grown until the OD₆₀₀ was 0.4–0.5. TDP-43 was then induced with 1 mM IPTG for 16 h at 15 °C. The *E. coli* cells were then lysed by sonication on ice in 50 mM HEPES (pH 7.5), 2% Triton X-100, 500 mM NaCl, 30 mM imidazole, 5% glycerol, 2 mM β -mercaptoethanol and protease inhibitors (cOmplete, EDTA-free, Roche). TDP-43 was purified over Ni-NTA agarose beads (Qiagen) and eluted from the beads using 50 mM HEPES (pH 7.5), 500 mM NaCl, 300 mM imidazole, 5% glycerol and 5 mM DTT. The protein was subsequently buffer-exchanged into 50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol and 5 mM DTT, flash-frozen in liquid nitrogen and stored as aliquots at –80 °C until use. Protein concentrations were determined by Bradford assay (Bio-Rad). The purity of TDP-43 was confirmed on a 4–20% polyacrylamide gel.

Thioflavin-T incorporation. Myo-granules were isolated from myotubes and diluted in PBS. Three separate biological replicates were performed constituting purification from three separate myotube cultures. First, 25 μ M thioflavin-T (Abcam) was added to recombinant 15 μ M HIS-SUMO–TDP-43, myo-granules or myo-granules plus recombinant 15 μ M HIS-SUMO–TDP-43. Subsequently, surface denaturation was performed with continuous shaking at 37 °C and thioflavin-T incorporation was then monitored every 10 min at 495 nm after excitation at 438 nm on a Gen5 microplate reader (BioTek). Finally, raw fluorescence values that were obtained for experimental conditions were background subtracted and plotted as a function of time. The resulting curves were fit to following a single exponential rate equation using Kaleidagraph (Synergy Software):

$$-Ae^{(-k_{\text{obs}}t)} + B \quad (1)$$

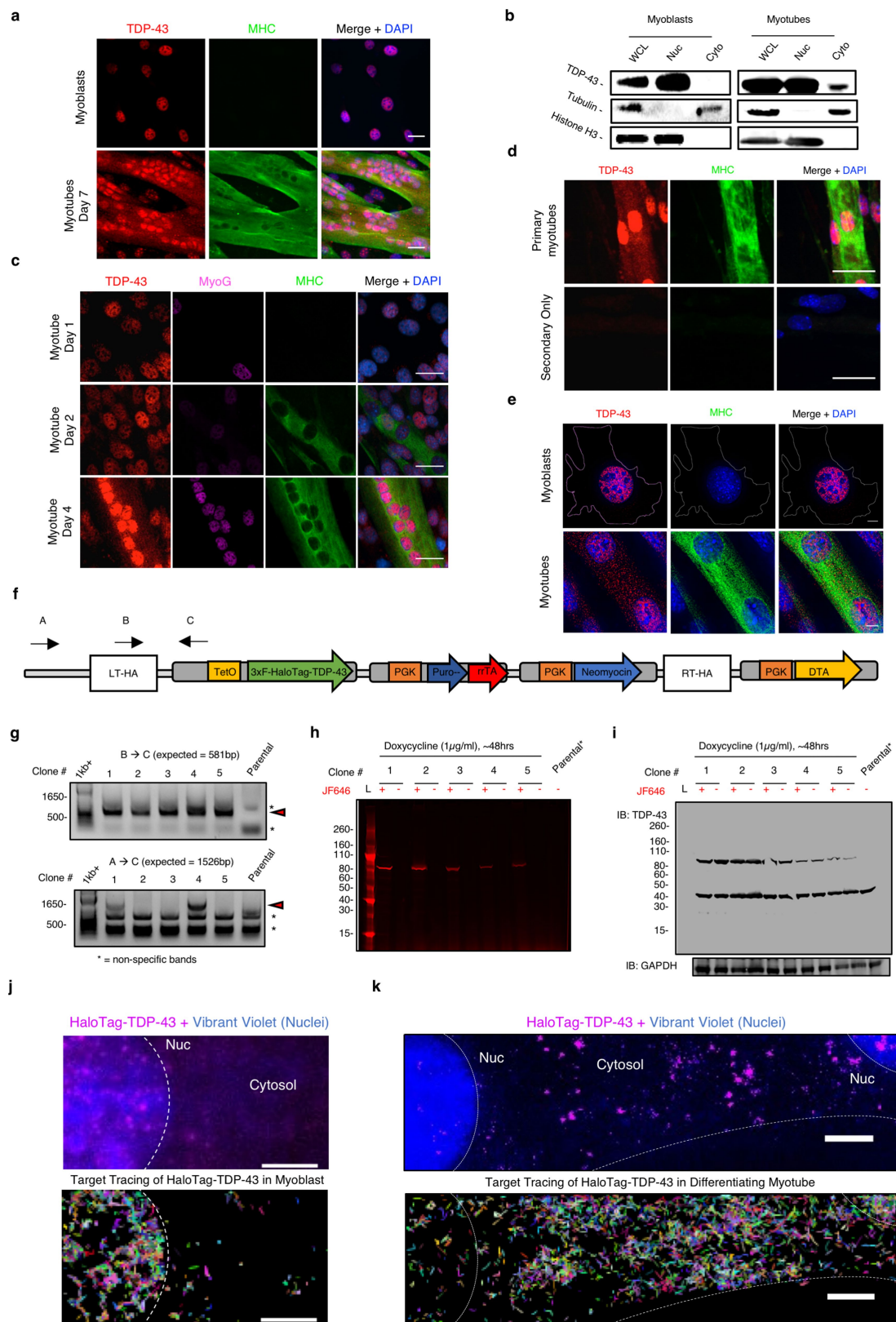
in which *A* is the amplitude, *k*_{obs} (min^{–1}) is a single exponential rate constant and *B* represents the maximal amount of fluorescence detected.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

eCLIP data are available from the Gene Expression Omnibus (GEO) under accession number GSE104796. Source Data are provided for Figs. 1d, 2c, 3f, h, 4c, 5b, c and Extended Data Figs. 1b, 2c, 3i–k, 4g, 7c, f, h, i, 8c, 9c. All other data supporting the findings of this study are available in the Supplementary Information. Data are available upon request from the corresponding authors.

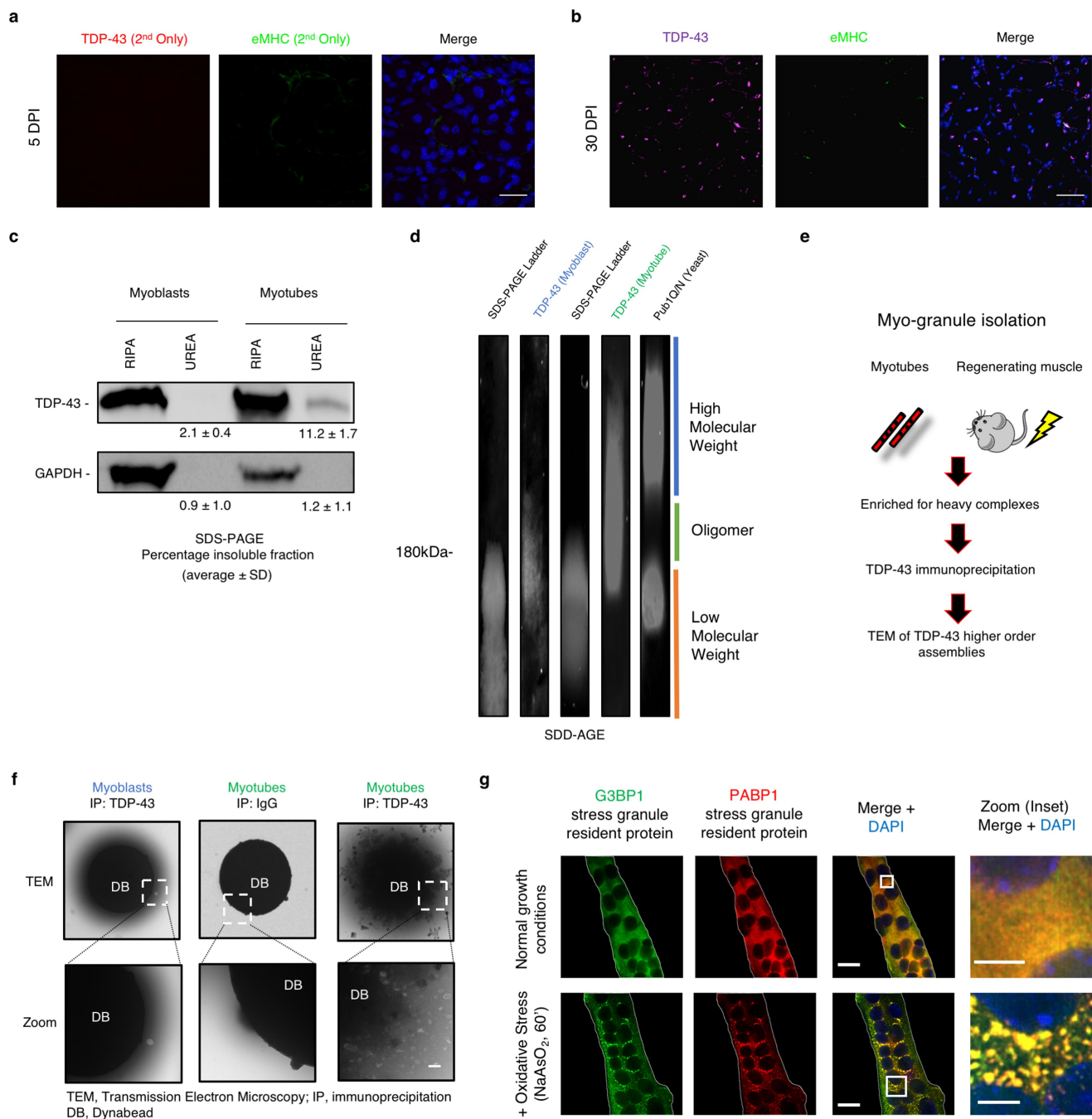
- Platt, R. J. et al. CRISPR–Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
- Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
- Halfmann, R. & Lindquist, S. Screening for amyloid aggregation by semi-denaturing detergent–agarose gel electrophoresis. *J. Vis. Exp.* **17**, 838 (2008).
- Fang, Y.-S. et al. Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun.* **5**, 4824 (2014).
- Jain, S. et al. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* **164**, 487–498 (2016).
- Winey, M., Meehl, J. B., O’Toole, E. T. & Giddings, T. H. Jr. Conventional transmission electron microscopy. *Mol. Biol. Cell* **25**, 319–323 (2014).
- Lovci, M. T. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* **20**, 1434–1442 (2013).
- Arnault, S., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5**, 687–694 (2008).



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Increased cytosolic TDP-43 during normal skeletal muscle formation. Related to Fig. 1. **a**, Nuclear localization of TDP-43 immunofluorescence in C2C12 myoblasts and both nuclear and cytoplasmic localization in C2C12 myotubes differentiated for seven days ($n = 3$ independent experiment). Myosin heavy chain (MHC) identifies differentiated cells. Scale bars, 25 μm . **b**, Subcellular fractionation reveals increased cytosolic TDP-43 in differentiating myotubes. Cytosolic (Cyto) myoblasts, $5.0 \pm 2.1\%$; cytosolic myotubes, $19.7 \pm 3.1\%$; $n = 3$ biologically independent experiments that showed similar results, unpaired, two-tailed Student's t -test, $P = 2.0 \times 10^{-3}$. **c**, Time course of TDP-43 expression during skeletal-muscle differentiation. $n = 3$ independent experiments with similar results. Myogenin (MyoG) (magenta) and MHC (green) identify differentiated cells. Nuclei were counterstained with DAPI. Scale bars, 25 μm . **d**, Top, TDP-43 expression in primary myotubes derived from muscle stem cells that were differentiated in culture for four days. $n = 3$ independent experiments with similar results. Bottom, images for a secondary-antibody only control. Scale bars, 25 μm . **e**, Deconvolution microscopy of TDP-43 expression in C2C12 myotubes differentiated for five days. Scale bar, 5 μm . $n = 3$ independent experiments with similar results. **f**, CRISPR–Cas9-mediated genomic integration of tetracycline-

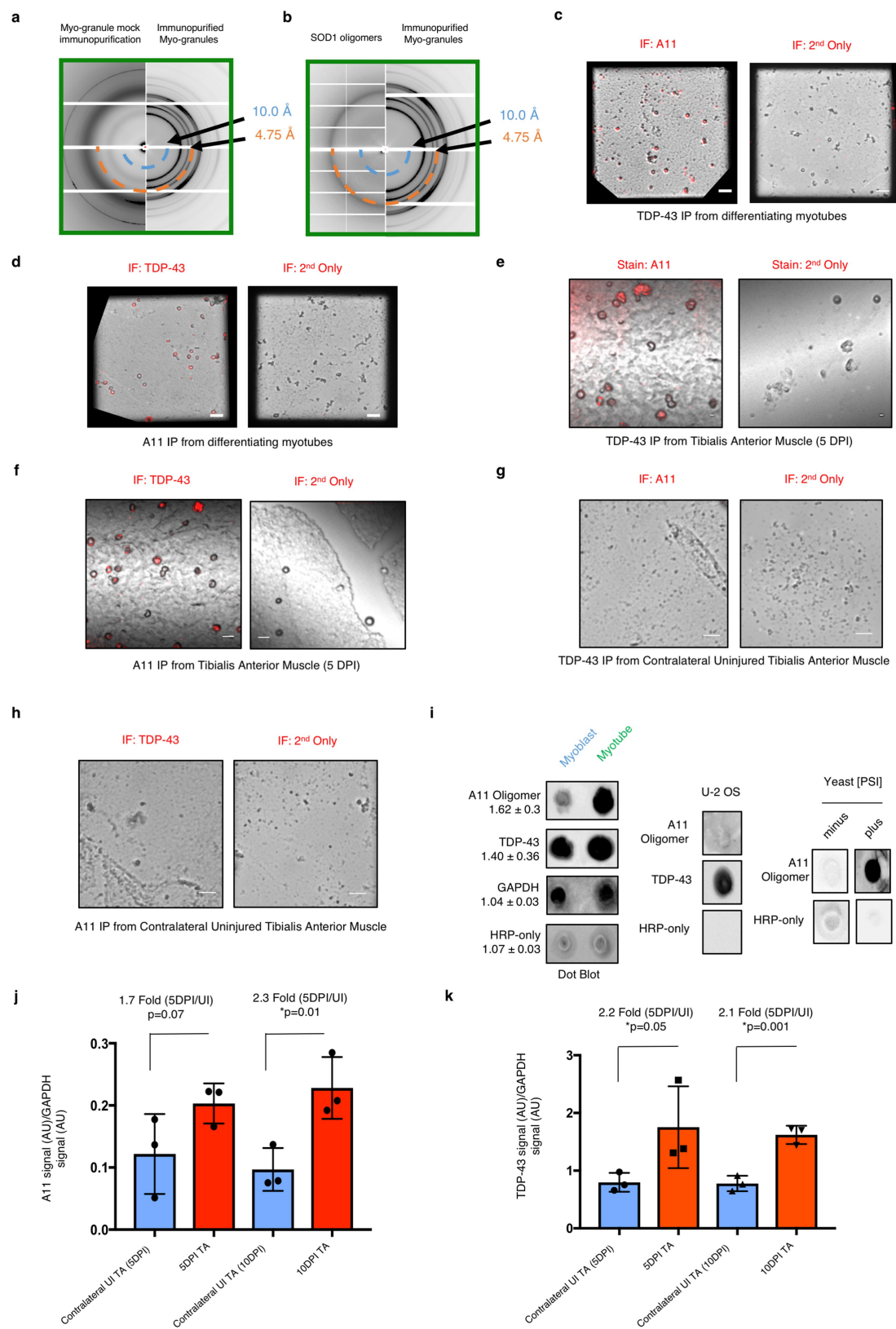
inducible HaloTag–TDP-43 into the *Rosa26* safe-harbour locus in C2C12 myoblasts. A, B and C represent approximate location of primers used in **g**. **g**, PCR analyses of gDNA from C2C12 myoblasts for the presence of the HaloTag–TDP-43 construct (top) and integration of the construct into the *Rosa26* locus (bottom) using the primers shown in **f**. $n = 3$ independent experiments with similar results. Red arrowheads point to the expected PCR product for integration of HaloTag–TDP-43 into *Rosa26*. Subsequent live-imaging experiments were performed using clones 1 and 4. Non-specific bands are indicated by an asterisk. **h**, Detection of fluorescently labelled HaloTag–TDP-43 in C2C12 myoblasts following induction resolved on SDS–PAGE. Janelia Fluor 646 (JF646). $n = 3$ independent experiments with similar results. **i**, Detection of both HaloTag–TDP-43 and endogenous TDP-43 in selected C2C12 cell clones. $n = 3$ independent experiments with similar results. **j**, **k**, Representative images of individual HaloTag–TDP-43 molecules in a myoblast (**j**) and a multinucleated myotube (**k**). Top, start of acquisition (frame 1). Nuclei (Nuc) and cytosolic borders are demarcated by white dotted lines. $n = 3$ independent experiments with similar results. Bottom, dynamic mapping of single TDP-43 molecule tracks using a multiple target tracing MATLAB script⁴⁶. Vibrant violet was used to detect myonuclei. Scale bars, 5 μm .



Extended Data Fig. 2 | During muscle formation TDP-43 adopts a higher-order state distinct from stress granules. Related to Fig. 1.

a, Secondary antibody-only control for TDP-43 staining of tibialis anterior muscle sections at 5 DPI. Scale bar, 25 μ m. $n = 5$ mice per condition, representative images are shown, all experiments showed similar results. Nuclei were counterstained with DAPI. **b**, Representative images of TDP-43 and eMHC immunostaining in tibialis anterior muscle sections at 30 DPI; nuclei were counterstained with DAPI. $n = 4$ mice. Scale bar, 50 μ m. **c**, RIPA-urea assay reveals the presence of an urea-insoluble TDP-43 fraction isolated from C2C12 myotubes that were differentiated for seven days, but not in C2C12 myoblasts. $n = 3$ independent experiments, each showing similar results, unpaired, two-tailed Student's t -test, $P = 0.0008$. GAPDH remains RIPA-soluble in both myoblasts and myotubes. $n = 3$ independent experiments, each showing similar results, unpaired, two-tailed Student's t -test, $P = 0.7443$. **d**, Higher molecular

weight SDS-resistant TDP-43 assemblies were present in differentiating C2C12 myotubes. Protein assemblies resolved by SDD-AGE. $n = 3$ independent experiments. Pub1Q/N-GFP from yeast forms SDS-resistant assemblies that have a higher molecular weight than TDP-43 assemblies. **e**, Schematic of the isolation of myo-granules that contain TDP-43 that are formed during skeletal muscle formation. **f**, Immunoprecipitation (IP) of TDP-43 on Dynabeads (DB) reveals that oligomers isolated from C2C12 myotubes are absent from myoblasts as observed by TEM. $n = 3$ independent experiments. **g**, Stress-granule formation in multinucleated myotubes derived from C2C12 cells. Immunofluorescence using antibodies against stress-granule proteins, G3BP1 and PABP1, after NaAsO₂ treatment or control conditions for 60 min. $n = 3$ independent experiments, each showing similar results. Zoom, boxed area shown at higher magnification. Scale bars, 5 μ m and 20 μ m (insets).



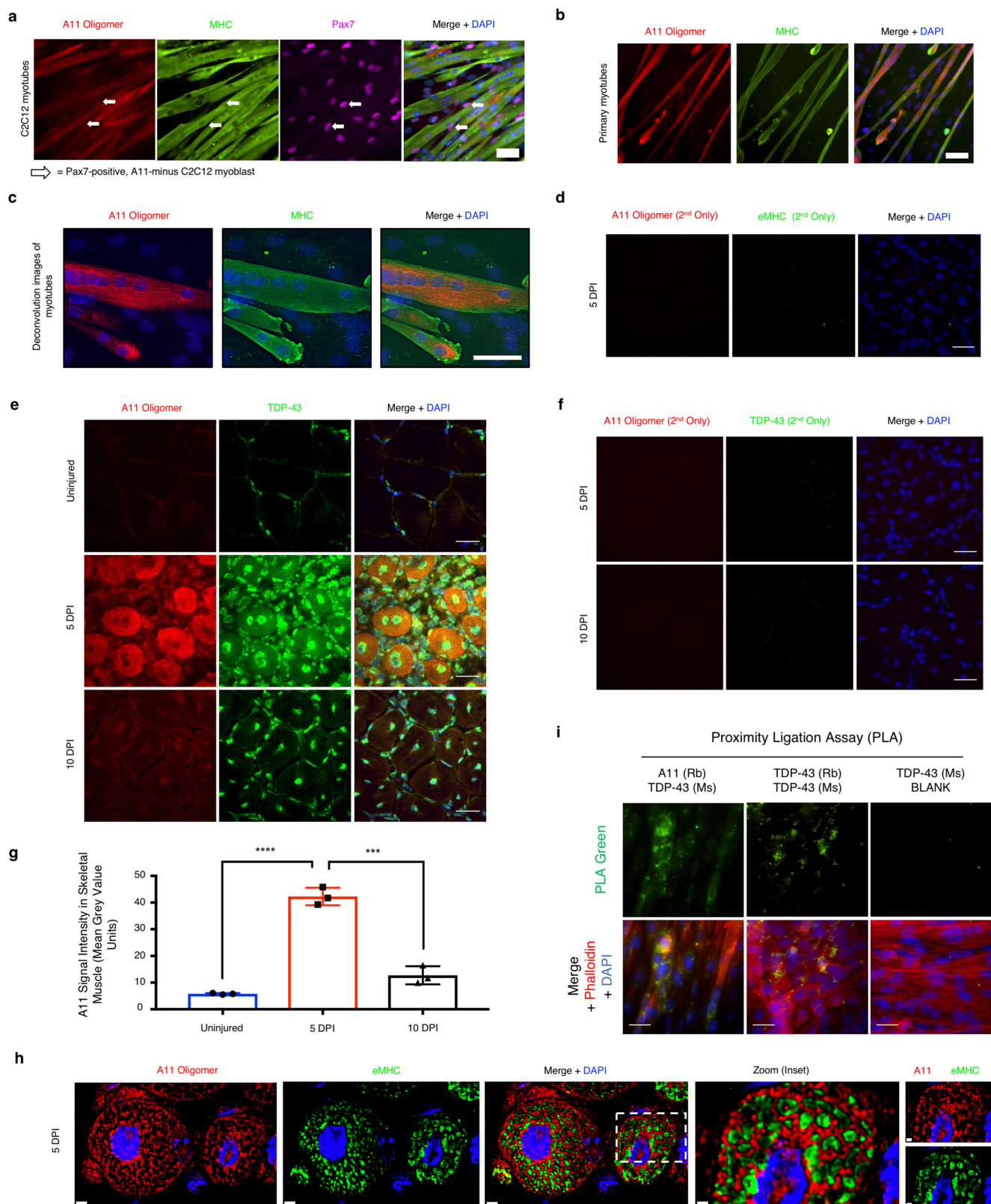
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Myo-granules isolated from cells and mice

contain TDP-43 and are amyloid-like oligomers. Related to Fig. 2.

a, b, X-ray diffraction of immunoprecipitated myo-granules (right half of **a, b**) compared to the diffraction of mock IgG immunoprecipitation (left half **a**) and to the diffraction of super oxide dismutase 1 (SOD1) amyloid oligomers (left half of **b**). For all diffraction patterns, two rings at approximately 4.8 Å and approximately 10 Å are drawn on the bottom half to highlight the absence of an approximately 4.8 Å reflection in the mock immunoprecipitation and a similar approximately 4.8 Å reflection with the absence of an approximately 10 Å reflection in the SOD1 diffraction. One sample per condition was used. Two diffraction images at different rotations were taken per sample and each image gave similar results. **c, d**, Complexes that were immunopurified using TDP-43 (**c**) or A11 (**d**) were isolated from C2C12 myotubes. Complexes express A11 (**c**) and TDP-43 (**d**), whereas immunopurified TDP-43 or A11 myo-granules that were immunostained with secondary antibodies only lack signal. Red, TDP-43 or A11 immunoreactivity. $n = 3$ independent experiments. Scale bars, 1 μm . **e, f**, Complexes that were immunopurified using TDP-43 (**e**) or A11 (**f**) were isolated from tibialis anterior muscle at 5 DPI.

Complexes express A11 (**e**) and TDP-43 (**f**), whereas immunopurified TDP-43 or A11 myo-granules immunostained with secondary antibodies only lack signal. Red, TDP-43 or A11 immunoreactivity. $n = 3$ mice. Scale bars, 0.05 μm . **g**, TDP-43 immunopurified complexes isolated from an uninjured tibialis anterior muscle (contralateral to the 5 DPI muscle) reveal no complexes with an A11 oligomeric confirmation. $n = 3$ mice. Scale bars, 0.05 μm . **h**, A11 immunopurified complexes from an uninjured tibialis anterior muscle (contralateral to the 5 DPI muscle) reveal no complexes containing TDP-43. $n = 3$ mice. Scale bars, 0.05 μm . **i**, Dot blot of A11 immunoreactivity in C2C12 cells differentiated into myotubes compared to myoblasts. Quantification reflects fold change in dot blot signal from myoblast to myotube. Data are mean \pm s.d., $n = 3$ independent experiments. **j, k**, Quantification of the dot blot signal for A11 conformation complexes (**j**) and TDP-43 conformation complexes (**k**) during skeletal muscle regeneration at 5 DPI and 10 DPI compared to contralateral uninjured tibialis anterior muscle and normalized to the HRP-only signal. Quantification reflects fold change in dot blot signal. Data are mean \pm s.d., $n = 3$ mice, P values were obtained using unpaired, two-tailed Student's t -tests.



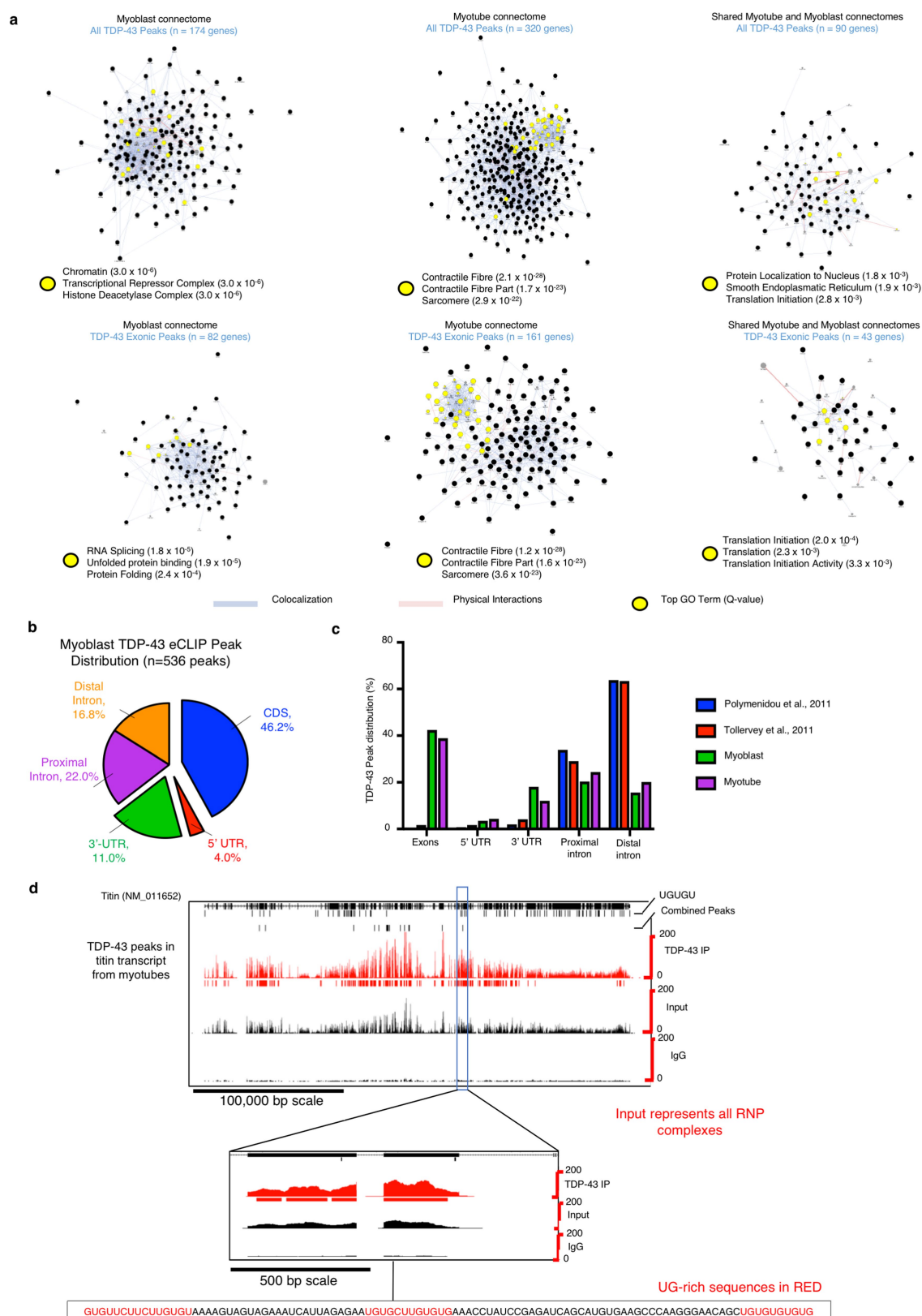
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Myo-granules in skeletal muscle contain TDP-43 and are amyloid-like oligomers. Related to Fig. 2. **a**, C2C12 myotubes differentiated for seven days reveal strong A11 immunoreactivity in MHC⁺ myotubes, but no A11 immunoreactivity in undifferentiated PAX7⁺ myoblasts. $n = 3$ independent experiments. Scale bar, 50 μm . **b**, Muscle stem cells isolated from four-month-old C57/BL6 mice were differentiated in culture for five days and show cytoplasmic and nuclear expression of A11 oligomers. Myotubes express MHC. Scale bar, 50 μm . $n = 3$ mice. **c**, Deconvolution microscopy of C2C12 myotubes differentiated for seven days reveal punctate A11 staining in MHC⁺ myotubes, but no A11 signal was found in undifferentiated myoblasts. Scale bar, 25 μm . $n = 3$ independent experiments. **d**, Secondary antibody-only control for A11 staining in tibialis anterior muscle sections at 5 DPI. Nuclei were counterstained with DAPI. Scale bar, 25 μm . $n = 4$ mice. **e**, Representative images of A11 and TDP-43 co-localization in tibialis anterior muscle for uninjured muscles, and at 5 DPI and 10 DPI. Scale bars, 25 μm . $n = 3$ mice. **f**, Secondary antibody-only control for A11–TDP-43

co-localization in tibialis anterior muscle sections at 5 DPI and 10 DPI shows a lack of signal. Nuclei were counterstained with DAPI. Scale bar, 25 μm . $n = 3$ mice. **g**, Quantification of A11 signal intensity in myofibres from **e**. Unpaired, two-tailed Student's *t*-test; comparison between uninjured muscle and 5 DPI, **** $P = 4.4 \times 10^{-5}$; comparison between 5 DPI and 10 DPI, *** $P = 4.1 \times 10^{-4}$; comparison between 10 DPI and uninjured muscle $P = 0.024$ (P value not shown). $n = 3$ mice per condition, $n = 10$ myofibres were averaged per mouse. Data are mean \pm s.d. **h**, Representative deconvolution image of A11 immunoreactivity and eMHC expression in the mouse tibialis anterior myofibres at 5 DPI that were quantified in Fig. 2c. $n = 3$ mice, each showing similar results. Scale bars, 2 μm and 0.8 μm (inset). **i**, Proximity ligation assays reveal complexes of TDP-43 and A11 (green) in C2C12 myotubes counterstained with phalloidin (red). A PLA positive control with two antibodies that recognize different epitopes of TDP-43 are positive, whereas complexes are absent if one primary antibody is omitted. $n = 3$ independent experiments per condition. Ms, mouse; Rb, rabbit.

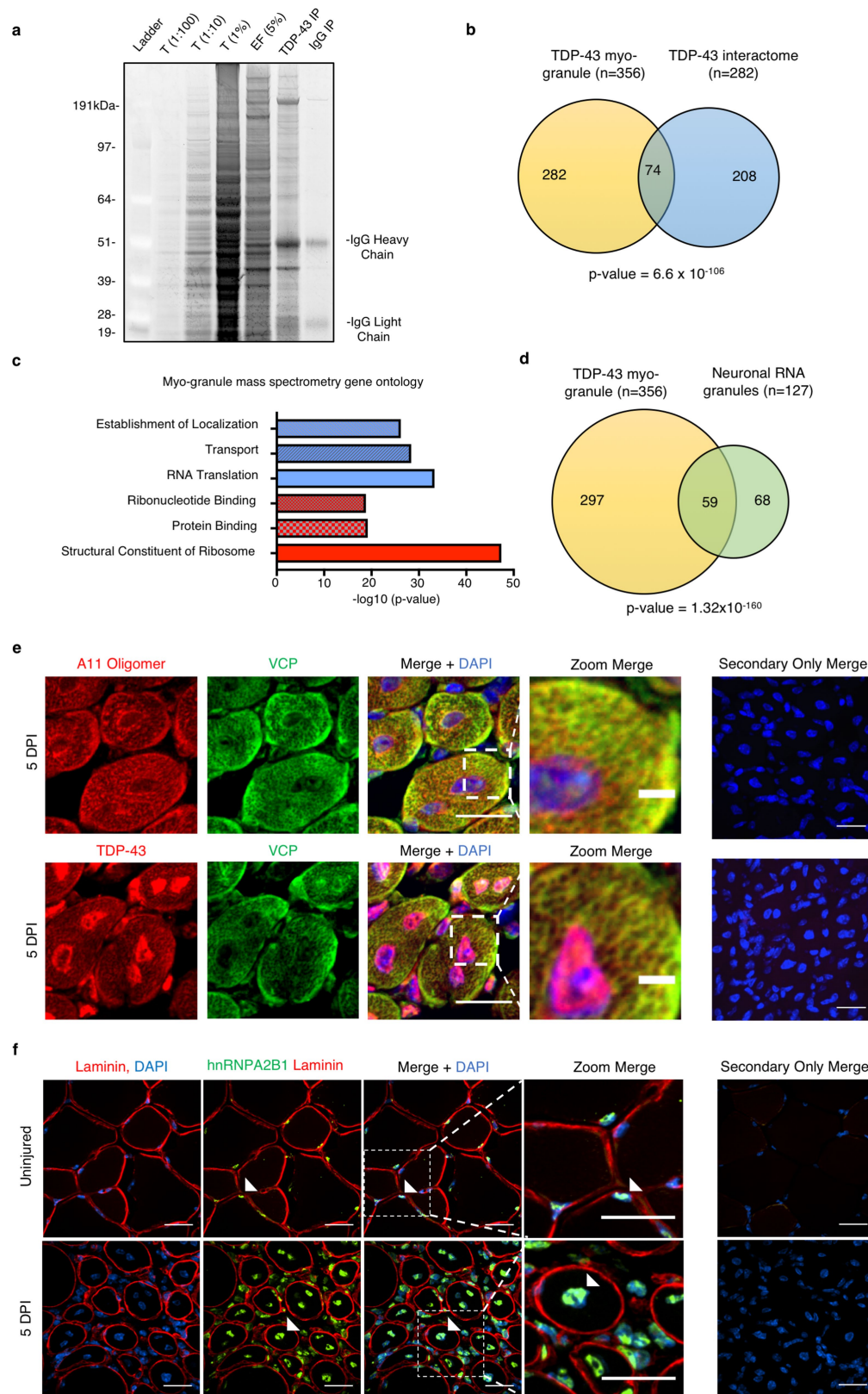
Extended Data Fig. 5 | TDP-43 eCLIP on skeletal muscle myoblasts and myotubes. Related to Fig. 3. **a**, RNA immunoprecipitation of C2C12 myotubes, followed by oligo-dT northern blot. Analyses reveal that A11 and TDP-43 associate with poly-A RNA. $n = 3$ biologically independent samples. **b**, Schematic of the eCLIP protocol for cultured C2C12 myoblasts and myotubes. **c**, Immunoprecipitation of TDP-43 complexes used for eCLIP in C2C12 myoblasts. $n = 2$ biologically independent samples. **d**, Same as in **c**, but for C2C12 myotubes. $n = 2$ biologically independent samples. **e**, Autoradiogram of ^{32}P -labelled TDP-43–RNA complexes fractionated by PAGE. White boxes indicate the area cut and used for eCLIP library preparation. $n = 1$ library was prepared per condition. **f**, Top, scatter plots indicate correlation between significant TDP-43 eCLIP peaks in biological replicates. Scatter plots represent fold enrichment for each region in TDP-43 eCLIP relative to paired size-matched input

with significant peaks in red ($P \leq 10^{-8}$ over size-matched input). P values for each peak to determine significance were calculated by Yates' χ^2 test (Perl), or Fisher exact test (R computing software) when the expected or observed read number was below five¹⁶. For myoblasts, R values were calculated using $n = 511,137$ non-significant peaks and $n = 596$ significant peaks. For myotubes, R values were calculated using $n = 413,368$ non-significant peaks and $n = 1,501$ significant peaks. Bottom, the UG-rich motif is significantly enriched in clusters from open reading frames and untranslated regions (UTRs). E values were determined using the DREME software tool. **g**, Irreproducible discovery rate analysis comparing peak fold enrichment across indicated datasets. **h**, TDP-43 eCLIP reveals that TDP-43 binds to the 3' UTR of the TDP-43 transcript in myoblasts (top) and myotubes (bottom). $n = 3$ biologically independent experiments, each showing similar results.



Extended Data Fig. 6 | TDP-43 binds to mRNAs that encode sarcomeric proteins during muscle formation. Related to Fig. 3. **a**, Myoblast (left), myotube (middle) and shared (right) connectome analysis for all TDP-43 eCLIP peaks (top) and TDP-43 exonic peaks (bottom). **b**, TDP-43 binds predominantly to exons of protein-coding RNAs in C2C12 myoblasts. **c**, Peak distribution for significantly enriched TDP-43 peak locations in

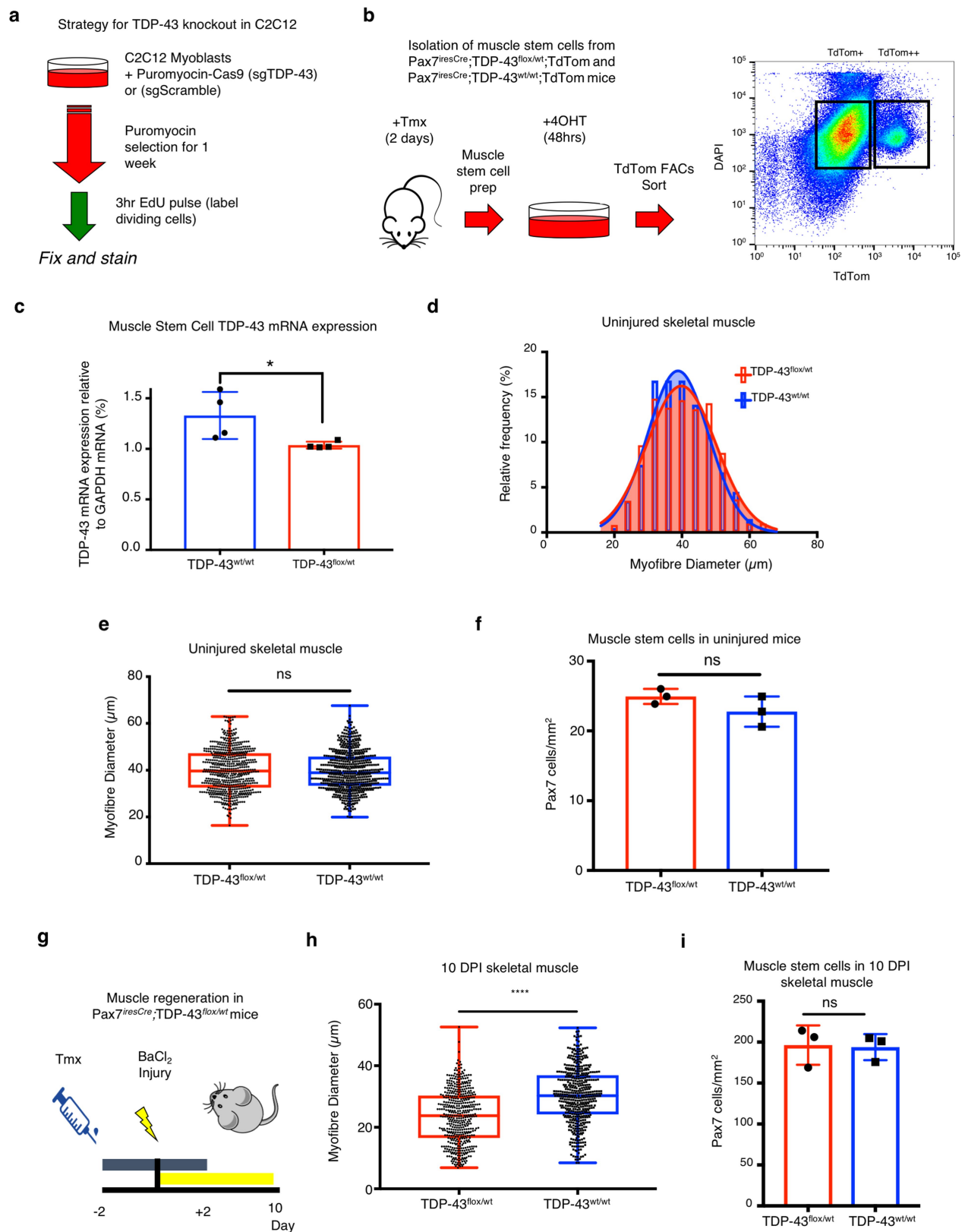
myoblasts and myotubes across the transcriptome reveal increased exonic and 3'-UTR associations compared to previously identified neuronal TDP-43 peaks^{18,19}. **d**, Identification of multiple TDP-43-binding sites across and within exons of *Ttn*. The zoomed region is representative of multiple UG-rich sequences within a single exon. $n = 3$ biologically independent experiments, each showing similar results.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Myo-granule protein composition. Related to Fig. 3. **a**, SDS–PAGE gel stained with SYPRO Ruby reveals enrichment of select proteins during fractionation of total cell lysate (T) from C2C12 myotubes, the enriched fraction (EF) and immunoprecipitation of TDP-43. $n = 3$ biologically independent experiments, each showing similar results. TDP-43 and IgG control immunoprecipitation experiments are representative of the fractions used for mass spectrometry. **b**, Venn diagram showing significant overlap between the myo-granule proteome and TDP-43 interactome (previously defined²²). The P value was determined using a hypergeometric test. **c**, Gene Ontology of myo-granules reveals enrichment for processes relating to the localization and

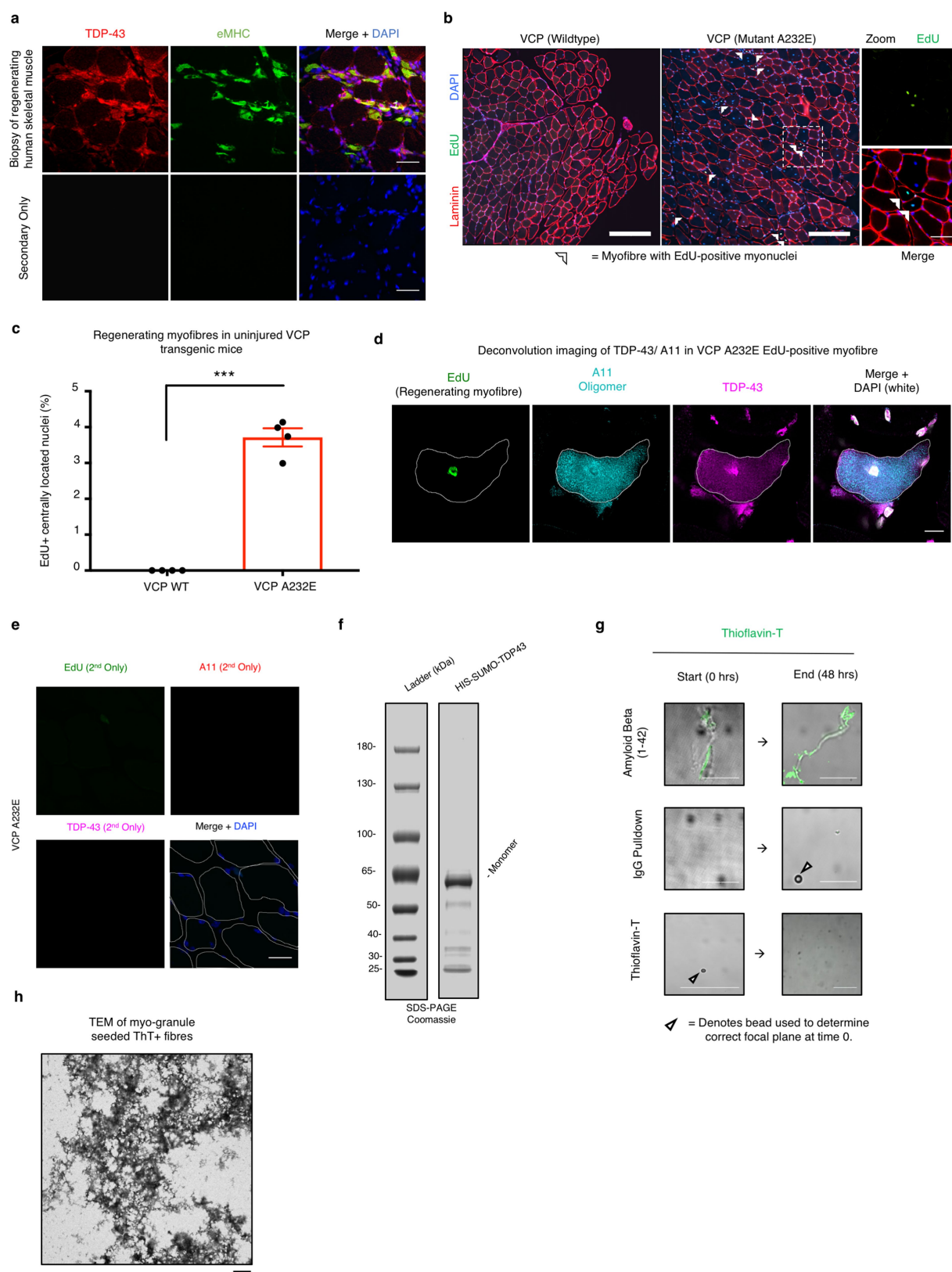
translation of RNA. $n = 356$ proteins, P values were determined using hypergeometric tests with Benjamini–Hochberg false-discovery rate corrections. **d**, Venn diagram showing significant overlap between myo-granules and neuronal RNA granule proteomes (previously defined²³). P value was determined using a hypergeometric test. **e**, VCP, a top hit in the myo-granule proteome, co-localizes with the cytoplasmic TDP-43 and A11 signals in mouse skeletal muscle at 5 DPI. $n = 3$ mice. **f**, The RNA-binding protein HNRNPA2B1 is not associated with the myo-granule proteome and remains localized in myonuclei in injured (5 DPI) and uninjured tibialis anterior muscle. $n = 3$ mice.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | TDP-43 is an essential protein for skeletal muscle formation. Related to Fig. 3. **a**, Schematic of the approach used to knockout *Tardbp* and quantify C2C12 myoblast proliferation. **b**, Schematic of the isolation and fluorescence-activated cell sorting (FACS) of muscle stem cells from *Pax7^{IREScree}Tardbp^{flox/WT}Rosa26^{tdTomato}* and *Pax7^{IREScree}Tardbp^{WT/WT}Rosa26^{tdTomato}* mice. More than 125,000 muscle stem cells were collected per mouse from two populations defined in **b** as TdTom⁺ and TdTom⁺⁺. **c**, *Tardbp* mRNA expression relative to *Gapdh* mRNA expression from isolated muscle stem cells from **b**. $n = 4$ independent experiments, each a mean of technical triplicates, from $n = 2$ mice. Unpaired, two-tailed Student's *t*-test, $*P = 0.0469$. Data are mean \pm s.d. **d**, Myofibre feret diameter frequency distribution in uninjured *Pax7^{IREScree}Tardbp^{flox/WT}* mice compared to *Pax7^{IREScree}Tardbp^{WT/WT}* controls. $n = 3$ mice, 600 myofibres were quantified per condition. **e**, Quantification of myofibre feret diameter shown in **c**. In the box plots, the horizontal bars show the mean, boxes show the 25th and 75th percentiles, whiskers show the minimum and maximum, individual

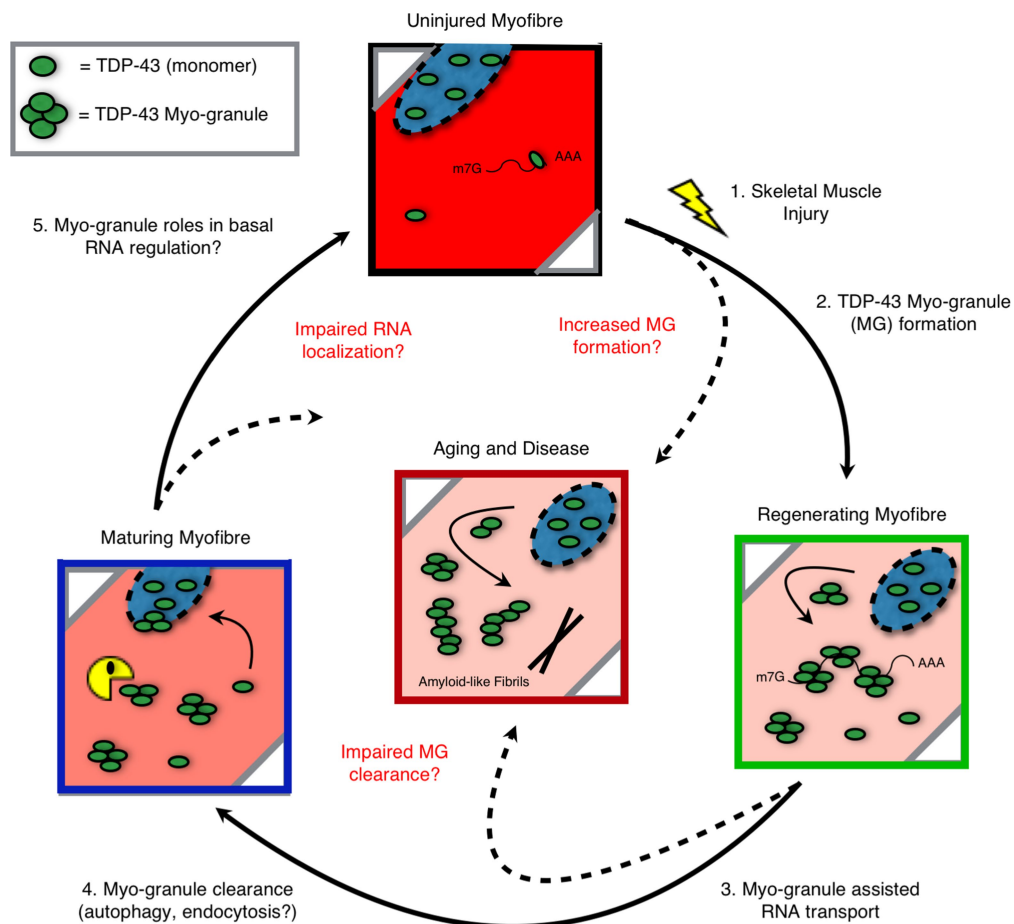
myofibres are shown as dots. $n = 3$ mice, 600 myofibres per condition. Unpaired, two-tailed Student's *t*-test, $P = 0.5925$; ns, not significant. **f**, *Pax7⁺* muscle stem cell numbers in uninjured *Pax7^{IREScree}Tardbp^{flox/WT}* mice compared to *Pax7^{IREScree}Tardbp^{WT/WT}* controls. $n = 3$ mice. Unpaired, two-tailed Student's *t*-test, $P = 0.1963$. Data are mean \pm s.d. **g**, Schematic of TDP-43 depletion in *Pax7⁺* muscle stem cells during muscle regeneration in *Pax7^{IREScree}Tardbp^{flox/WT}* and *Pax7^{IREScree}Tardbp^{WT/WT}* mice. Tmx, tamoxifen. **h**, Quantification of myofibre feret diameters from Fig. 3h at 10 DPI in muscle stem cells from *Pax7^{IREScree}Tardbp^{flox/WT}* mice compared to wild-type controls. In the box plots, the horizontal bars show the mean, boxes show the 25th and 75th percentiles, whiskers show the minimum and maximum, individual myofibres are shown as dots. $n = 489$ myofibres from $n = 3$ mice per condition. Unpaired, two-tailed Student's *t*-test, $***P = 2.3 \times 10^{-30}$. **i**, Similar *Pax7⁺* muscle stem cell numbers at 10 DPI in muscle stem cells from *Pax7^{IREScree}Tardbp^{flox/WT}* haploinsufficient mice compared to wild-type controls. Data are mean \pm s.d. from $n = 3$ mice. Unpaired, two-tailed Student's *t*-test, $P = 0.89$.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Myo-granules that seed amyloid-like fibres are increased in human muscle regeneration and in multisystem proteinopathy. Related to Figs. 4, 5. **a**, Representative images of TDP-43 expression (top) and secondary antibody-only control (bottom) in regenerating human skeletal muscle from a patient with necrotizing myopathy. $n = 3$ independent patient biopsies, each showing similar results. Scale bars, 50 μm . **b**, Representative tibialis anterior cross-section images of uninjured VCP(A232E) and VCP(WT) mice labelled with EdU after 21 days of EdU treatment in the drinking water to mark division and fusion of muscle stem cells. Laminin identifies myofibres and cells are stained with DAPI to identify nuclei. Arrowheads indicate myofibres with EdU⁺ centrally located myonuclei. $n = 3$ mice, each showing similar results. Scale bars, 200 μm and 50 μm (inset). **c**, Quantification of myofibres with EdU⁺ centrally located myonuclei in VCP(A232E) and VCP(WT) mice. $n = 4$ mice, over 1,000 myofibres were quantified per genotype. Data are mean \pm s.d. Unpaired, two-tailed Student's *t*-test,

$P = 6.5 \times 10^{-6}$. **d**, Representative deconvolution image of A11 and TDP-43 co-localization in a regenerating myofibre from a VCP(A232E) tibialis anterior muscle. $n = 3$ mice, each showing similar results. Scale bar, 10 μm . **e**, Secondary antibody-only control of uninjured VCP(A232E) tibialis anterior muscle sections reveals a lack of signal. Nuclei were counterstained with DAPI and myofibres were outlined in white. $n = 4$ mice, each showing similar results. Scale bar, 25 μm . **f**, Coomassie-stained recombinant HIS-SUMO-TDP-43 used for thioflavin-T assays resolved by SDS-PAGE. $n = 3$ biologically independent experiments, each showing similar results. **g**, Thioflavin-T incorporation reveals thioflavin-T⁺ amyloid-like fibres for recombinant amyloid- β_{1-42} and absence of thioflavin-T signal in both the IgG pull-down control and thioflavin-T alone. $n = 3$ biologically independent experiments, each showing similar results. Scale bars, 10 μm . **h**, Representative TEM image (zoomed out from Fig. 5e) of thioflavin-T⁺ (ThT) fibres formed from isolated myo-granules. $n = 3$ biologically independent experiments. Scale bar, 1 μm .



Extended Data Fig. 10 | Myo-granules in normal skeletal muscle regeneration and in disease. Schematic of TDP-43 oligomerization and aggregation in wild-type, ageing and diseased skeletal muscle myofibres.

Mechanoresponsive stem cells acquire neural crest fate in jaw regeneration

Ryan C. Ransom^{1,2,8}, Ava C. Carter^{3,8}, Ankit Salhotra^{1,2}, Tripp Leavitt¹, Owen Marecic^{1,2}, Matthew P. Murphy¹, Michael L. Lopez¹, Yuning Wei³, Clement D. Marshall¹, Ethan Z. Shen¹, Ruth Ellen Jones¹, Amnon Sharir⁴, Ophir D. Klein^{4,5,6}, Charles K. F. Chan^{1,2}, Derrick C. Wan¹, Howard Y. Chang^{3,7*} & Michael T. Longaker^{1,2*}

During both embryonic development and adult tissue regeneration, changes in chromatin structure driven by master transcription factors lead to stimulus-responsive transcriptional programs. A thorough understanding of how stem cells in the skeleton interpret mechanical stimuli and enact regeneration would shed light on how forces are transduced to the nucleus in regenerative processes. Here we develop a genetically dissectible mouse model of mandibular distraction osteogenesis—which is a process that is used in humans to correct an undersized lower jaw that involves surgically separating the jaw bone, which elicits new bone growth in the gap. We use this model to show that regions of newly formed bone are clonally derived from stem cells that reside in the skeleton. Using chromatin and transcriptional profiling, we show that these stem-cell populations gain activity within the focal adhesion kinase (FAK) signalling pathway, and that inhibiting FAK abolishes new bone formation. Mechanotransduction via FAK in skeletal stem cells during distraction activates a gene-regulatory program and retrotransposons that are normally active in primitive neural crest cells, from which skeletal stem cells arise during development. This reversion to a developmental state underlies the robust tissue growth that facilitates stem-cell-based regeneration of adult skeletal tissue.

The facial skeleton exhibits morphological variations that underlie the evolutionary diversification of mammals. The lower jaw comprises mandibular bone, vasculature, dentition, innervation and musculature. Mechanical forces are integral to skeletal homeostasis and skeletal regeneration by defining tissue architecture and driving cell differentiation. In the lower jaw, the mechanical forces applied during distraction osteogenesis promote endogenous bone formation across a mechanically controlled environment, providing functional replacement of tissue^{1,2}. Distraction osteogenesis has revolutionized the treatment of facial malformations that include Pierre-Robin sequence, Treacher Collins syndrome and craniofacial microsomia^{3–5}.

However, little is known about the cell population and molecular signals that drive tissue growth in distraction osteogenesis. Recently, the mouse skeletal stem cell (SSC) lineage has been elucidated and isolated⁶. Whether this lineage is present in the facial skeleton, which is known to arise from the neural crest, is unknown.

During regenerative processes, adult stem-cell populations change not only in proliferation and location but also in their underlying gene-regulatory programs^{7,8}. Stem cells may reactivate a greater potential for differentiation, while also responding to injury conditions⁹. Clinical studies comparing acute separation of bone to gradual distraction indicate that the application of constant physical force has a role in driving regeneration at the molecular level^{1–5}. The process of converting mechanical stimuli into a molecular response (mechanotransduction) occurs through multiple pathways, including the FAK pathway, leading to context-dependent transcriptional regulation¹⁰. Understanding how SSCs translate mechanical stimuli into productive regeneration will shed light on how force is transduced in other regenerative processes.

Here we use a rigorous model of mandibular distraction osteogenesis in mice and show that new bone is clonally derived from mandibular SSCs. Using the assay for transposase-accessible chromatin (ATAC-seq), as well as RNA sequencing (RNA-seq) to analyse the SSC transcriptome, we show that SSCs have distinct chromatin accessibility and gene expression within the FAK pathway. Activation of FAK through controlled mechanical advancement of the lower jaw in adults is required to induce a primitive neural crest transcriptional network that may allow for the massive tissue regeneration seen in distraction osteogenesis. The cellular mode of regeneration in response to mandibular distraction is of great interest, as this represents a successful strategy to elicit the endogenous potential of postnatal tissue^{11,12}.

Bone regeneration in distraction osteogenesis

We interrogated the cellular and mechanical mechanisms of adult bone regeneration by developing a mouse model of mandibular distraction osteogenesis, beginning with the design and three-dimensional (3D) printing of distraction devices (Fig. 1a, b). Next, animals were divided into four groups (Extended Data Fig. 1a): sham-operated (in which the mandible was exposed and the distraction device was placed, but there was no surgical cutting of the bone (osteotomy)); fracture (osteotomy without distraction); acutely lengthened (osteotomy with bone segments separated to 3 mm on day 5); and gradually distracted (osteotomy with bone segments separated by 0.15 mm every 12 hours, to a total separation of 3 mm).

Microcomputed tomography (μ CT) and pentachrome staining revealed that sham-operated mandibles had normal bone morphology and minimal new bone at postoperative day (POD) 43 (Fig. 1c). Fractured mandibles demonstrated incomplete bone union with persistent cartilage at the middle and end of the consolidation phase (Fig. 1d).

¹Department of Surgery, Division of Plastic and Reconstructive Surgery, Stanford University School of Medicine, Stanford, CA, USA. ²Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. ³Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA. ⁴Department of Orofacial Sciences and Program in Craniofacial Biology, University of California, San Francisco, CA, USA. ⁵The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, USA. ⁶Department of Pediatrics and Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. ⁷Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. ⁸These authors contributed equally: Ryan C. Ransom, Ava C. Carter. *e-mail: howchang@stanford.edu; longaker@stanford.edu

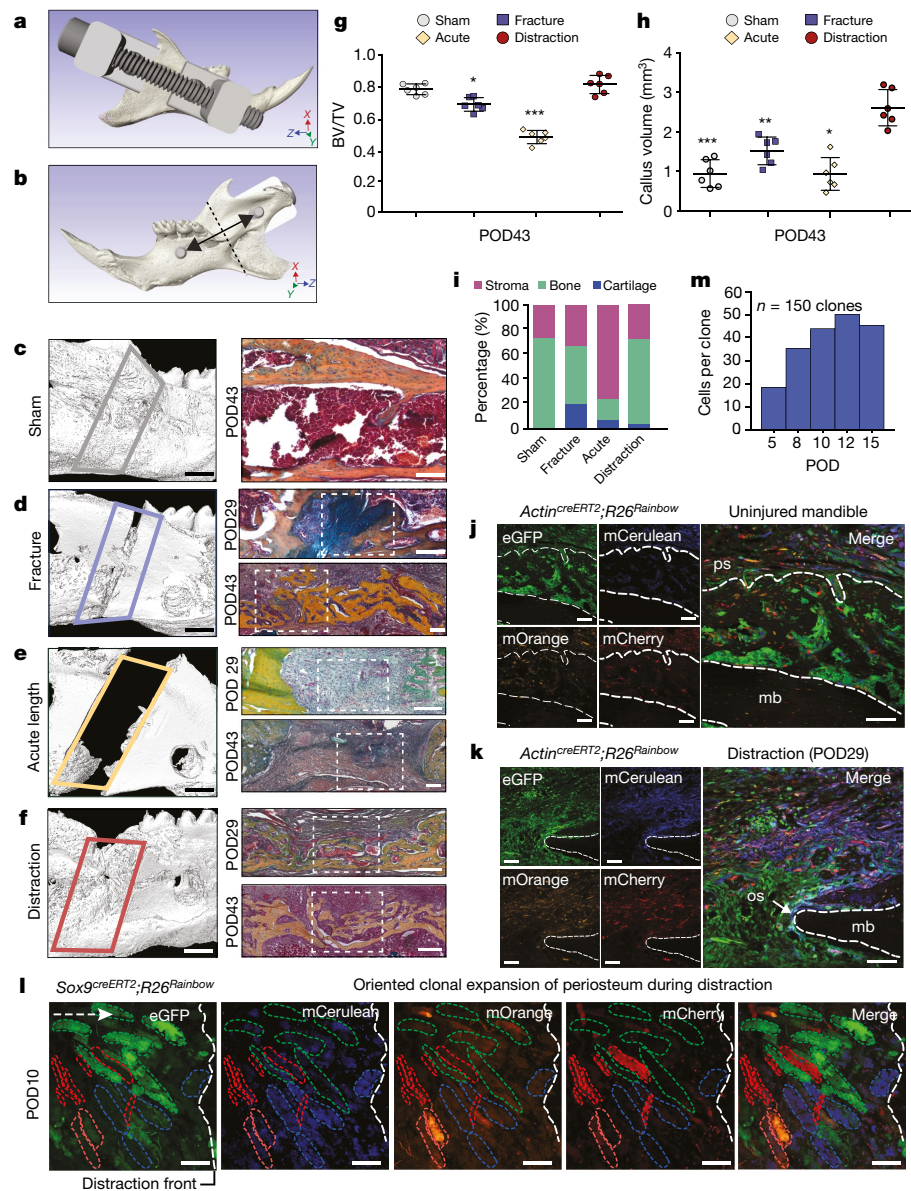


Fig. 1 | Tissue-resident stem and progenitor cells enact bone regeneration in distraction osteogenesis. **a**, Computer-assisted design of a distraction device using 3D μ CT of the C57BL/6 mouse hemimandible. **b**, The lingual aspect illustrates the location of the osteotomy (dotted line), perpendicular to the vector of bidirectional distraction (solid arrow). **c**, Three-dimensional μ CT of a sham-operated mandible (left, lateral view), with pentachrome staining of a transverse section (right) at POD43. The outlined area (left) indicates the volume analysed for new bone formation ($n = 5$). **d**, As for **c**, but for a fractured mandible, and also showing POD29. The white dotted lines indicate the osteotomy area. **e**, As for **d**, but for an acutely lengthened mandible. **f**, As for **d**, but for a gradually distracted mandible. **g**, Quantification of bone volume/total tissue volume (BV/TV) analysed at POD43 from μ CT ($n = 6$, $*P \leq 0.05$, $***P \leq 0.001$; Tukey's multiple comparisons). **h**, Quantification of new bone callus volume formed at POD43 ($n = 6$, $*P \leq 0.05$, $***P \leq 0.001$, $***P \leq 0.001$; Tukey's multiple comparisons). **i**, Tissue fraction of bone,

cartilage and stroma from pentachrome histology at POD43 ($n = 6$). **j**, Confocal micrographs of uninjured transverse mandible sections after one week ($n = 4$). Uninjured mandibular bone (mb) is between the dotted lines. eGFP, enhanced green fluorescent protein; mCerulean, membrane cerulean (blue fluorescence signal); mCherry, membrane cherry (red fluorescence signal); mOrange, membrane orange (orange fluorescence signal); ps, periosteum. **k**, As for **j**, but for POD29 of distraction osteogenesis ($n = 4$). The dotted outline indicates the mandibular bone at the distraction site. os, osteotomy site. **l**, Confocal micrographs of whole-mount periosteum at POD10 (clones are indicated by coloured dotted outlines) ($n = 5$). The vector of distraction is indicated by the dotted white arrow (left). Buccal to lingual view of periosteal callus overlying the distraction sites. **m**, Quantification of average clone size within the regenerate during distraction (POD5–15; $n = 5$). Scale bars, 1 mm (c–f left), 200 μ m (c–f right, j–l).

Acute lengthening resulted in non-union with primarily fibrous tissue (Fig. 1e). By contrast, gradual distraction resulted in complete union and robust new bone formation at mid- and end-consolidation (Fig. 1f). The callus mineralized volume fraction (bone volume/total tissue volume, or BV/TV) was significantly increased in gradual-distraction specimens compared with acutely lengthened ($***P < 0.001$) and fracture ($*P < 0.05$) specimens at POD43 (Fig. 1g). The callus volume (TV) of gradual-distraction specimens was significantly higher

than in all other conditions, including acute lengthening ($*P < 0.05$), fracture ($**P < 0.01$) and sham ($***P < 0.001$) (Fig. 1h). Analysis of tissue fraction in distraction and sham mandibles revealed similar proportions of mineralized bone, and confirmed cartilaginous healing in fracture specimens and fibrous healing in acute lengthening specimens (Fig. 1i). Whereas gradual-distraction mandibles exhibited direct intramembranous ossification, fracture mandibles displayed endochondral ossification (Extended Data Fig. 1b).

We next sought to investigate the cellular mechanism responsible for regeneration in distraction osteogenesis. We first ruled out the possibility of a circulating source of regenerated tissue using parabiosis (Extended Data Fig. 1c–f). Then, to test whether the generation of new mandible in distraction osteogenesis involved tissue-resident stem cells (Extended Data Fig. 1c), we performed the procedures on the mandibles of ubiquitous Rainbow mice (*Actin^{creERT2};Rosa26^{Rainbow}*; *Rosa26* is hereafter referred to *R26* and *Actin* refers to *Actb* throughout). After recombination of the Rainbow reporter (*R26^{VT2/GK3}*), cells are genetically marked with one of ten colour combinations, which is passed to all daughter cells (Extended Data Fig. 1g). To determine the location of stem and progenitor cells within the mandible, we traced uninjured tissue over the course of one year, finding large single-coloured clones within the periosteum—demonstrating the presence of a stem and progenitor population (Extended Data Fig. 1h).

We then sought to determine the role of these cells during distraction. *Actin^{creERT2};R26^{Rainbow}* mice were distracted, using uninjured mandibles for comparison. At mid-consolidation, we observed expansion of single-coloured clones near the osteotomy site (Fig. 1j, k). To trace the fate of cells within the periosteum during distraction osteogenesis, we developed a strategy of local labelling and found that large clone sizes were enriched (Extended Data Fig. 1i).

To assess the lineage-specific characteristics of regeneration in distraction, we distracted mandibles from skeleton-specific *Sox9^{creERT2};R26^{mT/mG}* mice and found that cells of the *Sox9*-expressing lineage gave rise to new bone (Extended Data Fig. 2a–e). We then crossed the *Sox9^{creERT2}* driver to the Rainbow construct and observed infiltration of multiple single-coloured clones into the distraction gap at early and late distraction (PODs 8 and 12) (Extended Data Fig. 2f, g). Clonal analysis of the regenerate over time revealed that cells within the lateral periosteum expanded into large clones with linear alignment (Fig. 1l and Extended Data Fig. 2h, i). As with *Actin^{creERT2};R26^{Rainbow}* mandibles, *Sox9^{creERT2};R26^{Rainbow}* mandibles exhibited equivalent recombination frequencies at POD12 for each fluorescent protein (Extended Data Fig. 2j–l). The average size of the *Sox9*-lineage clone increased over the course of distraction, peaking at POD12 (Fig. 1m). Thus, skeletal-lineage-specific stem or progenitor cells are responsible for new bone formation during distraction osteogenesis.

Next, we purified skeletogenic populations from mandibles using a strategy that has been used for long bones⁶. Lineage tracing of *Sox9^{creERT2};R26^{Rainbow}* mice for two weeks revealed clonal expansion within the growth plate of the mandibular condyle (Fig. 2a). We therefore collected tissues from the condylar growth plate to enrich for skeletal stem cells (SSCs, with markers $CD45^- Ter119^- CD202b^- Thy1^+ 6C3^- CD51^+ CD105^- CD200^+$), their multipotent progenitors (bone, cartilage and skeletal progenitors (BCSPs): $CD45^- Ter119^- CD202b^- Thy1^+ 6C3^- CD51^+ CD105^+$) and their unipotent progenitors (osteoprogenitors: $CD45^- Ter119^- CD202b^- Thy1^+ 6C3^- CD51^+ CD105^+ CD200^-$; and chondroprogenitors: $CD45^- Ter119^- CD202b^- Thy1^+ 6C3^- CD51^+ CD105^+ CD200^+$) (Extended Data Fig. 3a). We sorted single SSCs and demonstrated tertiary colony formation (Extended Data Fig. 3b). There was a complete absence of circulating (green fluorescent protein (GFP)⁺) SSCs and BCSPs, as assessed by parabiosis (Extended Data Fig. 3c–f).

Distraction osteogenesis elicited an expansion of SSCs and BCSPs that is not seen following fracture (Fig. 2b). Expansion after fracture was delayed and restricted to osteoprogenitors. Colony formation and proliferation were greatly elevated in SSCs (Fig. 2c–f and Extended Data Fig. 3g–i) at the time that distraction-specific kinetics were seen. Formation of mineralized tissue was substantially elevated by distraction (d-)SSCs versus fracture (f-)SSCs (Fig. 2e, f).

Distraction-specific gene regulation

We next investigated the molecular changes that drive the marked differences in regenerative capacity displayed by the SSC lineage in response to mechanotransduction. We performed ATAC-seq on SSCs, BCSPs and osteoprogenitors isolated from uninjured, fractured and

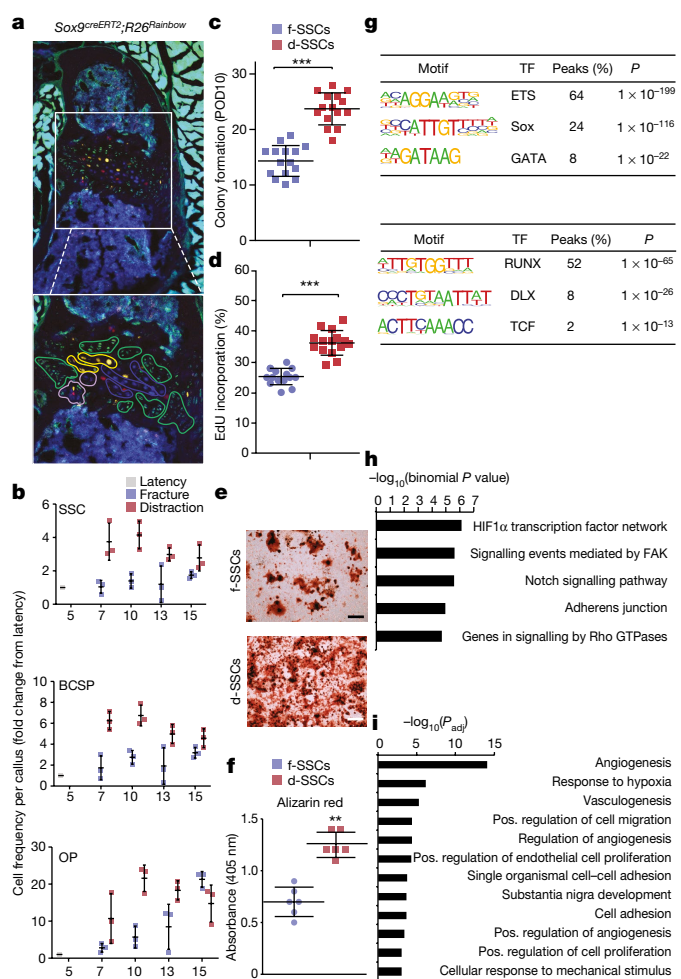


Fig. 2 | Transcriptional regulation underlying SSC function. **a**, Top, confocal micrograph of a transverse section from a *Sox9^{creERT2};R26^{Rainbow}* mandibular condyle ($n = 6$); bottom, expanded image of the area in the white box. Coloured outlines indicate single-colour clones. **b**, Quantification of cellular frequency within mandibular calluses in fracture and distraction conditions ($n = 3$ per time point; boundaries indicate standard deviation (s.d.)). **c**, Quantification of colonies formed from f- or d-SSCs isolated at POD10 ($n = 15$, $***P < 0.0001$; *t*-test). **d**, $n = 15$, $**P < 0.01$ and **e**, $n = 15$, $*P < 0.05$ SSCs demonstrated elevated colony-forming capacity compared with uninjured SSCs ($n = 6$). **d**, Quantification of proliferation rate via incorporation of 5-ethynyl-2'-deoxyuridine (EdU) into SSCs at POD10 ($n = 15$, $***P < 0.0001$; *t*-test). **e**, **f**, Alizarin red staining of f- versus d-SSCs ($n = 6$, $**P = 0.001$; *t*-test) at POD10. **g**, Motif analysis for differential peaks shared between SSCs and BCSPs. Top, chromatin sites that are more accessible in distraction than in fracture ($n = 1,617$ sites). Bottom, sites that are more accessible in fracture than in distraction ($n = 1,506$ sites). The top three motifs and the transcription factors (TFs) that bind them, the percentage of sites that contain the motif and the one-sided binomial *P*-values with Benjamini–Hochberg corrections are shown. **h**, GO terms, identified by ATAC-seq, enriched for genes near peaks that are more accessible in d- than in f-SSCs and BCSPs ($n = 1,617$ sites). Selected GO terms (right) with significant *P*-values (one-sided binomial, with Benjamini–Hochberg corrections; left) are shown. **i**, GO terms enriched for genes upregulated in d- versus f-SSCs at PODs 10 and 15, identified by RNA-seq. Select GO terms with significant *P*-values (one-sided binomial, with Benjamini–Hochberg corrections) are shown. Scale bars, 200 μ m (**a**, **e**).

distracted mandibles at POD10 to capture transcription-factor-binding sites throughout the genome¹³ (Extended Data Fig. 4a). Samples clustered into three groups: differentiated osteoprogenitors (d-OPs and f-OPs), quiescent cells (SSCs, BCSPs and f-SSCs), and activated or regenerative cells (d-SSCs, d-BCSPs and f-BCSPs) (Extended Data Fig. 4b, d).

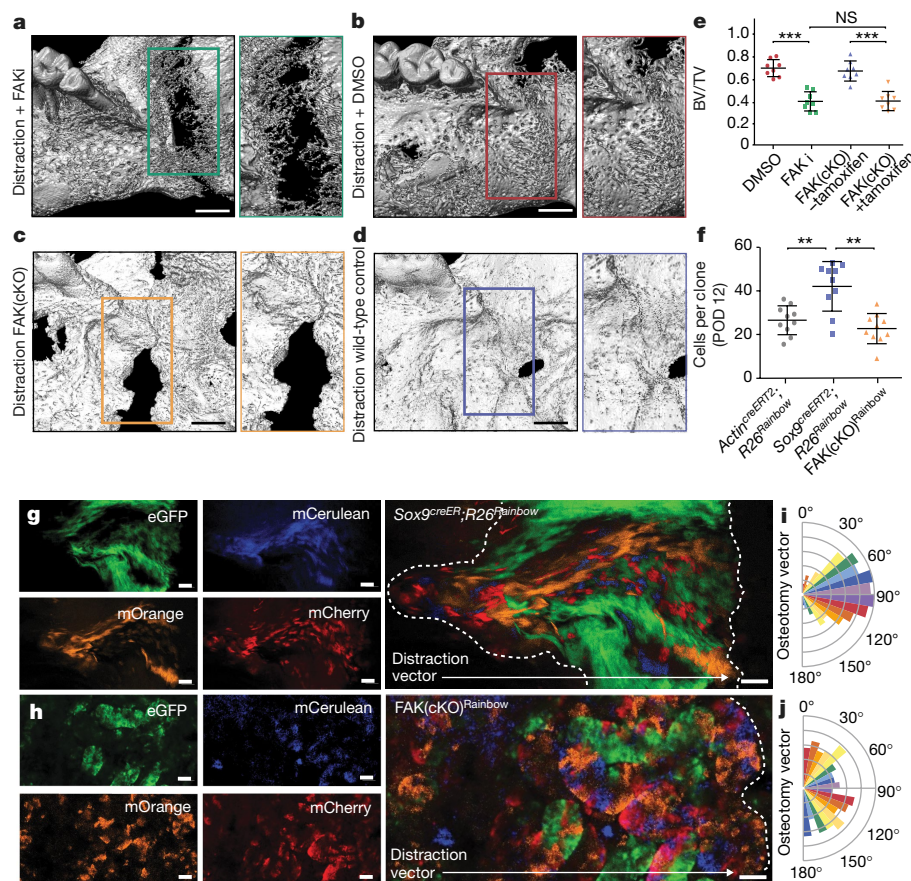


Fig. 3 | FAK inhibition disrupts bone formation during distraction. **a**, Three-dimensional μ CT of a distracted mandible treated with FAK inhibitor (FAKi) during distraction osteogenesis, collected at POD29. An absence of bone union is indicated by the green outline (with high magnification shown at the right). The view is a lingual aspect. $n = 5$. **b**, As for **a**, but using dimethylsulfoxide (DMSO) as a control in place of the FAK inhibitor. **c**, As for **a**, but using conditional FAK knockout (FAK(cKO)) animals treated with tamoxifen two weeks before distraction. **d**, As for **a**, but using wild-type animals and a corn-oil control treatment. **e**, BV/TV analysed at POD29 ($n = 8$, $***P \leq 0.001$; NS, not significant; Tukey's multiple comparisons). **f**, Quantification of average clone size within the regenerate over the course of distraction (POD12) in *Actin^{creERT2};R26^{Rainbow}*, *Sox9^{creERT2};R26^{Rainbow}*, and *FAK(cKO)^{Rainbow}* mice ($n = 10$ per condition; Tukey's multiple comparisons). **g**, Lateral to medial view of the callus overlying the distraction site. Confocal micrographs of whole-mount *Sox9^{creERT2};R26^{Rainbow}* periosteum at POD12 (the white dotted outline shows the distraction gap; $n = 6$). **h**, As for **g**, but for *FAK(cKO)^{Rainbow}* mandible. **i**, Quantification of angular clonal expansion in *Sox9^{creERT2};R26^{Rainbow}* mice at POD12. **j**, As for **i**, but for *FAK(cKO)^{Rainbow}* mice. Scale bars, 1 mm (**a–d**), 200 μ m (**g, h**).

We performed RNA-seq on d-SSCs, f-SSCs, d-BCSPs, f-BCSPs, d-OPs and f-OPs at PODs 5, 10 and 15, and demonstrated major differences between fracture and distraction osteogenesis at the level of SSCs and BCSPs, whereas the osteoprogenitors were highly similar in fracture and distraction conditions (Extended Data Fig. 4f, g). d-SSCs at PODs 10 and 15 were more similar to one another than d-SSCs were to f-SSCs at a single time point, suggesting that their transcriptional programs bifurcated according to the applied mechanical force. The expression of molecules from developmental signalling pathways and of functional molecules was highly variable between fracture and distraction osteogenesis, and only somewhat variable between time points (Extended Data Fig. 5a–c).

We directly compared ATAC-seq from d- and f-SSCs and d- and f-BCSPs. We found many changes in accessible chromatin sites (14,370 changes in SSCs; 9,208 in BCSPs), of which 3,123 were shared (Extended Data Fig. 4c, e). Motif analysis showed that sites that are less accessible in distraction osteogenesis than in fracture are enriched for binding motifs for the core transcription factors that drive skeletal development (RUNX and DLX) (Fig. 2g). *Runx2* and *Dlx5* were highly and differentially expressed and accessible between f- and d-SSCs, making them strong candidates for binding factors (Extended Data Fig. 4e and Extended Data Fig. 5d, e). Binding sites for these factors in pre-osteoblasts were less accessible in d-SSCs and d-BCSPs than in uninjured or fracture cells^{14,15} (Extended Data Fig. 5f, g).

Sites that gain accessibility during distraction osteogenesis are enriched for ETS, SOX and GATA motifs (Fig. 2g). Genes near distraction-specific sites are enriched for Gene Ontology (GO) terms associated with the hypoxia-inducible factor 1 α (HIF1 α), FAK, Notch and RAS signalling pathways, as well as with adherens junctions (Fig. 2h). At the RNA level, GO terms associated with vascularization, adherens junctions, cell migration and responses to mechanical stimuli are enriched in d-SSCs (Fig. 2i and Extended Data Fig. 4h–j). Genes upregulated in d-BCSPs are enriched for migration and adhesion terms, whereas genes upregulated in f-BCSPs are enriched for cartilage-formation terms (Extended Data Fig. 4k–m). Osteoprogenitors show fewer changes in expression in fracture versus distraction, indicating that most molecular changes during distraction osteogenesis occur in SSCs and BCSPs (Extended Data Fig. 4n–p).

We next explored the role of FAK signalling in distraction, because this pathway was upregulated in d-SSCs, and because FAK is a known mechanotransducer. FAK propagates information about physical stimuli at integrin-mediated cell–matrix contacts to the nucleus, affecting proliferation, differentiation, migration and more. The downstream targets of FAK are context-dependent and thus we sought to understand its role in our model^{9,10}.

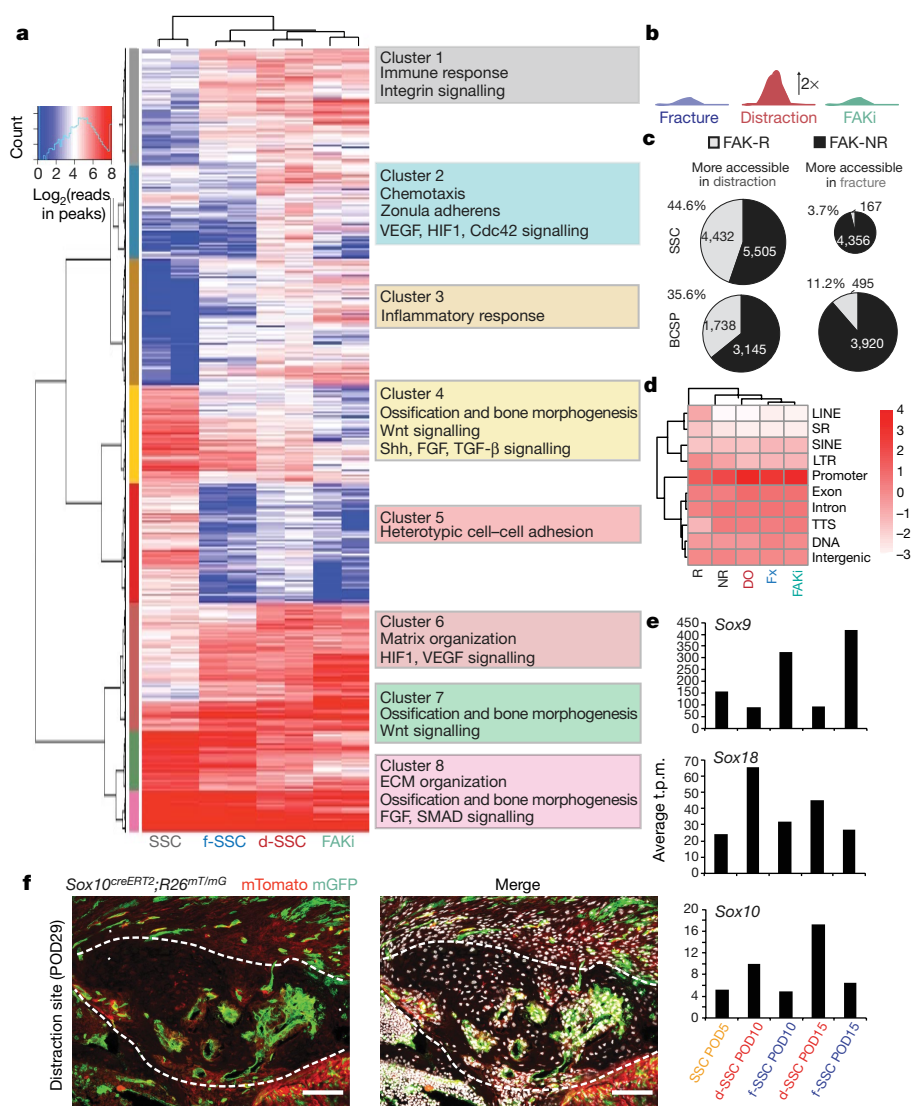


Fig. 4 | Changes in chromatin state during distraction osteogenesis with FAK inhibition. **a**, Heat map showing differential peaks (\log_2 (normalized reads in peaks)) between SSCs, f-SSCs and d-SSCs. *k*-means clusters are indicated by coloured bars at the left. Enriched GO terms for each cluster are shown at the right. **b**, Graphical demonstration of FAK-R sites. **c**, Pie charts showing the proportion of accessible sites that are FAK-R in SSCs and BCSPs. For SSCs, 4,432/9,937 sites (44.6%) that are more accessible in the distraction condition than in the fracture condition are FAK-R, versus 167/4,523 sites (3.7%) in the fracture condition ($P < 1 \times 10^{-16}$; one-sided Fisher's exact test). For BCSPs, 1,738/4,883 sites (35.6%) that are more accessible in distraction are FAK-R, versus 495/4,415 sites (11.2%) that are more accessible in the fracture condition ($P < 2.2 \times 10^{-16}$; one-sided Fisher's exact test). **d**, Genomic enrichment for peaks in five categories of

SSC: R, FAK-R; NR, FAK-NR; DO, distraction osteogenesis; Fx, fracture; and FAKi. **e**, Expression of *Sox9*, *Sox10* and *Sox18* as determined by RNA-seq in d- and f-SSCs from PODs 5, 10 and 15. The y-axis shows the average t.p.m. (transcripts per million) from two biological replicates. All genes are significantly differentially expressed ($P < 0.05$, from DESeq2; see Methods) between the fracture and distraction conditions at at least one time point. **f**, Confocal micrographs of a distraction regenerate in a *Sox10^{creERT2};R26^{mT/mG}* mandible collected at POD29 (the area of new bone is outlined with the dotted white lines). mTomato, background; mGFP, *Sox10*-expressing lineage; white, 4',6-diamidino-2-phenylindole (DAPI) staining. Representative of three independent experiments. Scale bar, 200 μ m.

Inhibiting FAK disrupts bone formation

To investigate whether mechanotransduction via FAK is essential in distraction osteogenesis, we inhibited FAK with the small molecule PF-573228 (FAKi condition). In addition, we evaluated a skeleton-specific conditional knockout of FAK (FAK(cKO)) using *Sox9^{creERT2};Ptk2^{fl/fl}* animals (Extended Data Fig. 6a). μ CT analysis revealed that bone formation was diminished in both FAKi and FAK(cKO) mandibles compared with controls (Fig. 3a–e; $***P < 0.001$). Pentachrome staining revealed cartilage formation under FAKi conditions that was not present in controls at PODs 15 and 29 (Extended Data Fig. 6b–e). Disruption of bone formation and induction of cartilage at the distraction site was a notable tissue-level response to FAK inhibition, congruent with disruption of the integrin-mediated

cell–matrix interactions that are responsible for intramembranous bone formation in distraction osteogenesis.

FAKi stunted skeletal lineage differentiation and progenitor expansion (Extended Data Fig. 6f–i). Ex vivo imaging and intracellular fluorescence-activated cell sorting (FACS) confirmed that FAKi inhibited FAK phosphorylation (Extended Data Fig. 6j). Similarly, induction of FAK(cKO) led to a disruption of the normal skeletal stem and progenitor expansion in distraction osteogenesis (Extended Data Fig. 7a, b). SSCs isolated from FAKi and FAK(cKO) mandibles at POD10 revealed significantly reduced colony-forming potential (Extended Data Fig. 6k; FAKi: $***P < 0.001$; FAK(cKO): $***P < 0.001$).

To further characterize the effect of FAK(cKO), we carried out in vivo clonal analysis using *Sox9^{creERT2};Ptk2^{fl/fl};R26^{Rainbow}* mice

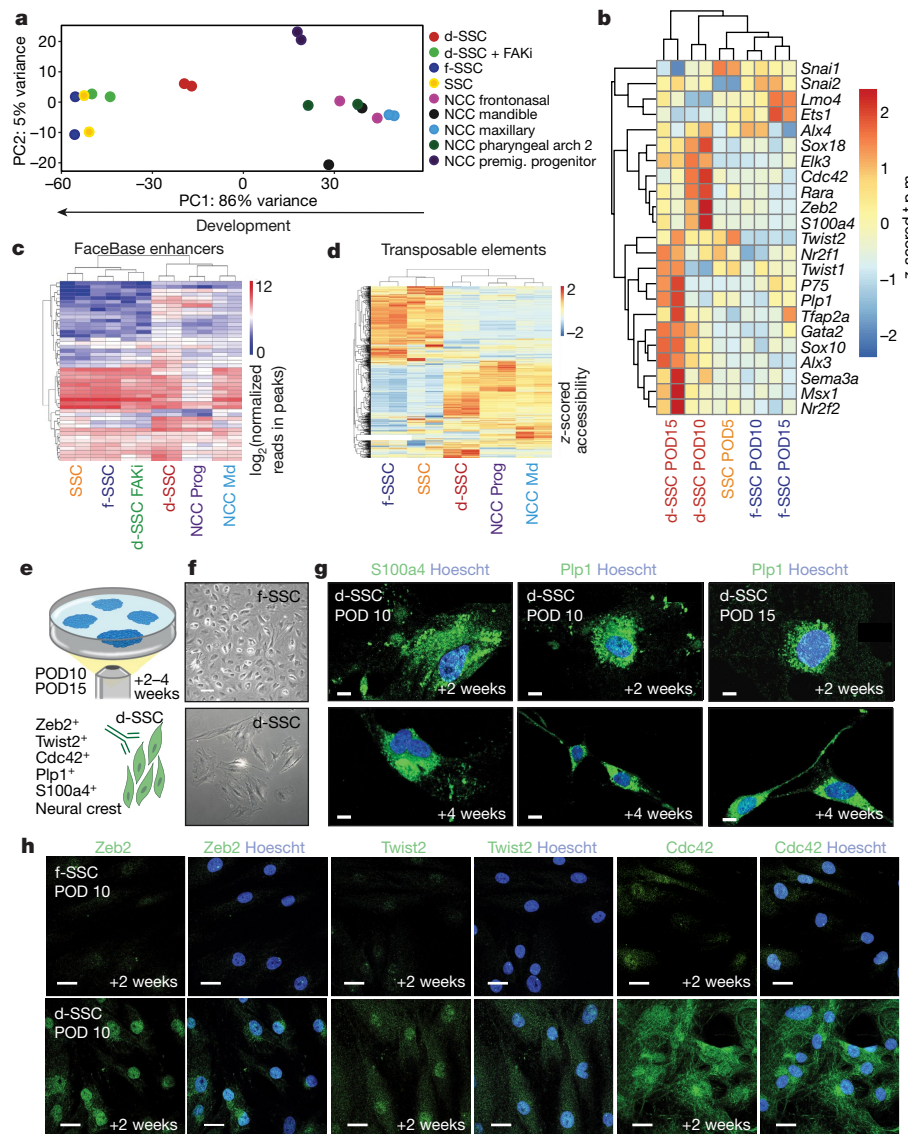


Fig. 5 | SSCs adopt an NCC-like state during distraction.

a, Principal component analysis (for principal components (PC) 1 and 2) of ATAC-seq data from SSCs and NCCs ($n = 269,900$ peaks). The arrow at the bottom shows the direction of cellular differentiation from NCC to SSC. **b**, Heat map of neural crest gene expression in d- and f-SSCs at PODs 5, 10 and 15. The z-scored t.p.m. refers to the number of standard deviations from the mean for each t.p.m. value in a sample. *P75* is also known as *Ngfr*. **c**, Accessibility at mouse neural crest enhancers (from FaceBase^{24,25}) at neural crest genes. Shown are $\log_2(\text{reads in peaks})$ for ATAC-seq peaks that overlap the annotated enhancers. **d**, Heat map of z-scored accessibility at all families of transposable elements from ATAC-seq data. **e**, Schematic showing f- or d-SSC culture and staining

from postoperative mandibles. FACS-isolated f-SSCs and d-SSCs at PODs 10 and 15 form colonies at two weeks and four weeks in culture (top). Then d- and f-SSCs (bottom right, green cells) are stained with the known NCC markers indicated at the bottom left (indicated by the green antigen). **f**, Representative images of d- or f-SSCs plated and cultured for two weeks. **g**, Immunofluorescent staining of POD10 and POD15 d-SSCs for NCC markers (*S100a4* and *Plp1*; green) at two and four weeks in culture. Images are representative of three independent experiments. Blue (Hoechst) staining identifies the nucleus. **h**, Immunofluorescent staining of f-SSCs versus d-SSCs at POD10 for the markers *Zeb2*, *Twist2* or *Cdc42* (green) and nuclear staining (blue) (representative of three independent experiments). Scale bars, 200 μm (f), 100 μm (g, h).

(hereafter FAK(cKO)^{Rainbow}) for distraction osteogenesis (Extended Data Fig. 7c, d). FAK(cKO)^{Rainbow} mice exhibited clones of significantly smaller size in the distraction site versus controls (Fig. 3f–h; $**P < 0.01$). Whereas a normal clonal recombination frequency was observed in *Sox9^{creERT2};R26^{Rainbow}* and FAK(cKO) mandibles during distraction (Extended Data Fig. 7e–h), FAK(cKO) disrupted the oriented clonal expansion observed in *Sox9^{creERT2};R26^{Rainbow}* mandibles (Fig. 3i, j), highlighting that integrin-based cell–matrix interactions are required for vector-aligned expansion of progenitors.

To understand the molecular role of FAK during distraction osteogenesis, we performed ATAC-seq on SSCs and BCSPs treated with FAKi, finding that many distraction-specific accessible chromatin sites were no longer accessible in d-SSCs + FAKi. In FAKi mandibles, d-SSCs and d-BCSPs have chromatin accessibility resembling that of

f-SSCs and BCSPs at a large subset of sites that differ between the fracture and distraction osteogenesis conditions (Fig. 4a; BCSP peaks not shown). *k*-means clustering divided all differential peaks between SSCs, f-SSCs and d-SSCs into eight clusters. We assigned GO terms to nearest genes. Injury-related terms were more accessible in fracture, distraction and distraction + FAKi samples (clusters 1, 3 and 6), whereas Wnt and fibroblast growth factor (FGF) signalling were enriched in clusters that were more accessible in the uninjured state (clusters 4, 7 and 8). We focused on distraction-specific sites the gain in accessibility of which was blocked with FAK inhibition, representing mechanotransduction-sensitive sites (clusters 2 and 5). These sites were enriched for adhesion and migration terms, along with terms for vascular/endothelial growth factor (VEGF), HIF1, and *Cdc42* signalling (Fig. 4a).

We separated sites that were differentially accessible between the fracture and distraction conditions into FAK-responsive (FAK-R; more or less accessible in distraction than in fracture and FAKi) or FAK-non-responsive (FAK-NR; more or less accessible in distraction and distraction plus FAKi than in fracture; Fig. 4b and Supplementary Table 1). The majority of FAK-R sites in SSCs and BCSPs increased in accessibility (44.6% and 35.6%, respectively) rather than decreased (3.7 and 11.2% respectively) compared with fracture (Fig. 4c), suggesting that FAK is required for activation of a large part of the distraction-specific regenerative program. Genomic annotation revealed that FAK-R sites were more enriched for transposable long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs) than were FAK-NR sites or all sites (Fig. 4d). Transposable elements constitute around 50% of the genome and contain transcription-factor-binding sites. Interestingly, LINEs are the largest class that contains active enhancers in the neural crest, from which the mandible is derived¹⁶.

To understand the mechanosensitive program further, we performed motif enrichment in FAK-R and FAK-NR sites. CCAAT/enhancer-binding protein (C/EBP) and activating transcription factor (ATF) motifs were enriched in FAK-NR sites in SSCs and BCSPs, consistent with the roles of C/EBP, ATF4 and RUNX2 in bone differentiation *in vitro*¹⁷. FAK-R sites in SSCs were enriched for Sox and ETS motifs, whereas responsive sites in BCSPs were enriched for Sox but not ETS motifs (Extended Data Fig. 8a). Sox9 is a core transcription factor in osteogenesis, and other Sox proteins are involved in mandibular development and other regenerative processes. RNA-seq revealed that Sox9 was downregulated in d-SSCs. Sox18 and Sox10 were upregulated, indicating that they may direct the FAK response (Fig. 4e and Extended Data Fig. 8b, d). The ETS factor Elk3 showed similar upregulation (Extended Data Fig. 8c). We were surprised to find Sox10, Sox18 and Elk3 upregulated in distraction osteogenesis, as they are not involved in postnatal bone repair, but are critical for neural crest development¹⁸.

Lineage tracing with Sox10^{CreERT2};R26^{mT/mG} mice confirmed Sox10 expression within newly forming bone (Fig. 4f and Extended Data Fig. 8e–g). Given the expression of neural crest transcription factors by d-SSCs and the migratory properties of these cells (Fig. 1m, n and Fig. 3g, h), we hypothesized that d-SSCs access a primitive, neural crest cell (NCC)-like state to achieve productive lengthening.

SSCs adopt an NCC-like state during distraction

NCCs are highly plastic embryonic cells that give rise to diverse tissues including the Schwann cells, teeth and bones of the mandible^{19–22}. To assess whether d-SSCs take on a more NCC-like identity, we compared their chromatin accessibility to that of premigratory mandibular NCCs (NCC Prog) and postmigratory populations (NCC Md (mandibular), Mx (maxillary), PA2 (pharyngeal arch 2) and FNP (frontonasal projection))²³ (Extended Data Fig. 9a). Principal component analysis (PCA) showed that SSCs, f-SSCs, and d-SSCs + FAKi cluster close together, whereas d-SSCs fall closer to NCCs (Fig. 5a). Clustering on all open chromatin sites, d-SSCs are more similar to NCC populations than are SSCs, f-SSCs or d-SSCs + FAKi (Extended Data Fig. 9b). We identified regulatory elements that are inaccessible in homeostatic SSCs but accessible in d-SSCs and NCCs, and found that they are enriched for embryonic, developmental and migratory terms (Extended Data Fig. 9c, d). Inaccessibility of these DNA elements in f-SSCs and d-SSCs + FAKi indicated that this activation is FAK-dependent. RNA-seq confirmed that key neural crest genes are expressed specifically in d-SSCs (Fig. 5b). d-SSCs gained accessibility at bona fide developmental neural crest enhancers (FaceBase^{24,25}) in a FAK-dependent manner (Fig. 5c). Profiling of SSC and NCC populations revealed a subset of FAK-R sites that are active in neural crest populations, further indicating FAK dependency (Extended Data Fig. 9e). The association of FAK-R sites with LINEs (Fig. 4d) led us to consider the relationship between this NCC program and transposable-element insertions. Transposable elements that were accessible in SSCs were different from those accessible in NCCs, and the d-SSCs clustered closely with NCCs (Fig. 5d).

We assessed neural crest potential in these cells by sorting SSCs from distraction versus fracture mandibles (Fig. 5e, f). d-SSCs stained positive for the NCC-associated markers S100a4 and Plp1 at two and four weeks of culture (Fig. 5g). Additionally, canonical neural crest markers were specifically expressed in the d-SSC condition, further supporting the idea that an NCC-like program is activated (Fig. 5h and Extended Data Fig. 9f).

Discussion

Here we have comprehensively modelled the process of bone regeneration in the jaw from the tissue to the chromatin level, to understand how controlled mechanical separation of bones leads to lengthening of the mandible (Extended Data Fig. 10). We have shown that the skeletal stem cell lineage gives rise to new bone in this paradigm, and can be isolated using the same markers as for long bone. The differences in chromatin accessibility and gene expression between the fracture and distraction conditions suggest that, while stem-cell proliferation and differentiation to produce restricted progenitors are important for this massive tissue regeneration, they are not the only factors at play. The downregulation of the canonical skeletal program during distraction osteogenesis suggested that the homeostatic program for tissue renewal is not sufficient for regeneration in distraction. We have shown that an alternative regenerative program activated in distraction osteogenesis is dependent on the FAK pathway, which has a known role in mechanotransduction. This FAK-dependent program is an embryonic NCC-like program that reverts d-SSCs to a more plastic, developmental state.

Postnatal regeneration is highly restricted in vertebrates. Newts use both dedifferentiation and activation of tissue-resident stem cells and progenitors, whereas the axolotl uses dedifferentiation followed by redifferentiation in the limb^{26–31}. Studies in mouse digits have demonstrated the formation of new bone with lineage-restricted origins^{11,12}. We have shown, however, that adult stem cells revert to a developmentally plastic state during regeneration of the jaw. Reversion to an embryonic-like NCC has also been seen during the initiation of melanoma tumours³². In the context of the massive tissue regeneration seen in distraction osteogenesis and the cellular hyperproliferation that occurs in cancer, a more potent stem-cell-like program may be required, whereas the more restricted cell (that is, the postnatal SSC in uninjured tissue) is sufficient during homeostatic renewal.

Retrotransposon sequences are co-opted as transcriptional enhancers during development, and related sequences make up large transcriptional networks^{33–37}. The accessibility of differing retrotransposon families between uninjured and fractured SSCs versus distraction SSCs suggests that a novel transcriptional network is active in this regenerative context, and may have a role during neural crest development. Understanding the relationship between retrotransposon activation and developmental mechanisms may unveil key processes underlying regenerative paradigms.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0650-9>.

Received: 23 November 2017; Accepted: 15 August 2018;

Published online 24 October 2018.

- Ilizarov, G. A. The tension-stress effect on the genesis and growth of tissues: part II. The influence of the rate and frequency of distraction. *Clin. Orthop. Relat. Res.* **239**, 263–285 (1989).
- Ilizarov, G. A. The tension-stress effect on the genesis and growth of tissues. Part I. The influence of stability of fixation and soft-tissue preservation. *Clin. Orthop. Relat. Res.* **238**, 249–281 (1989).
- Tahiri, Y., Viesel-Mathieu, A., Aldekhayel, S., Lee, J. & Gilardino, M. The effectiveness of mandibular distraction in improving airway obstruction in the pediatric population. *Plast. Reconstr. Surg.* **133**, 352e–359e (2014).
- McCarthy, J. G., Schreiber, J., Karp, N., Thorne, C. H. & Grayson, B. H. Lengthening the human mandible by gradual distraction. *Plast. Reconstr. Surg.* **89**, 1–8 (1992).

5. Khansa, I. et al. Airway and feeding outcomes of mandibular distraction, tongue-lip adhesion, and conservative management in Pierre Robin sequence: a prospective study. *Plast. Reconstr. Surg.* **139**, 975e–983e (2017).
6. Chan, C. K. et al. Identification and specification of the mouse skeletal stem cell. *Cell* **160**, 285–298 (2015).
7. Adam, R. C. et al. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370 (2015).
8. Ge, Y. et al. Stem cell lineage infidelity drives wound repair and cancer. *Cell* **169**, 636–650 (2017).
9. Frechin, M. et al. Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour. *Nature* **523**, 88–91 (2015).
10. Bell, S. & Terentjev, E. M. Focal adhesion kinase: the reversible molecular mechanosensor. *Biophys. J.* **112**, 2439–2450 (2017).
11. Rinkevich, Y., Lindau, P., Ueno, H., Longaker, M. T. & Weissman, I. L. Germ-layer and lineage-restricted stem/progenitors regenerate the mouse digit tip. *Nature* **476**, 409–413 (2011).
12. Lehoczyk, J. A., Robert, B. & Tabin, C. J. Mouse digit tip regeneration is mediated by fate-restricted progenitor cells. *Proc. Natl Acad. Sci. USA* **108**, 20609–20614 (2011).
13. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
14. Hojo, H., Ohba, S., He, X., Lai, L. P. & McMahon, A. P. Sp7/Osterix is restricted to bone-forming vertebrates where it acts as a Dlx co-factor in osteoblast specification. *Dev. Cell* **37**, 238–253 (2016).
15. Meyer, M. B., Benkusky, N. A. & Pike, J. W. The RUNX2 cistrome in osteoblasts: characterization, down-regulation following differentiation, and relationship to gene expression. *J. Biol. Chem.* **289**, 16016–16031 (2014).
16. Prescott, S. L. et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
17. Tominaga, H. et al. CCAAT/enhancer-binding protein β promotes osteoblast differentiation by enhancing Runx2 activity with ATF4. *Mol. Biol. Cell* **19**, 5373–5386 (2008).
18. Rogers, C. D., Phillips, J. L. & Bronner, M. E. Elk3 is essential for the progression from progenitor to definitive neural crest cell. *Dev. Biol.* **374**, 255–263 (2013).
19. Santagati, F. & Rijli, F. M. Cranial neural crest and the building of the vertebrate head. *Nat. Rev. Neurosci.* **4**, 806–818 (2003).
20. Kaukua, N. et al. Glial origin of mesenchymal stem cells in a tooth model system. *Nature* **513**, 551–554 (2014).
21. Kragl, M. et al. Analysis of neural crest-derived clones reveals novel aspects of facial development. *Sci. Adv.* **2**, e1600060 (2016).
22. Kaukua, M. et al. Oriented clonal cell dynamics enables accurate growth and shaping of vertebrate cartilage. *eLife* **6**, e25902 (2017).
23. Minoux, M. et al. Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. *Science* **355**, eaal2913 (2017).
24. Brinkley, J. F. et al. The FaceBase Consortium: a comprehensive resource for craniofacial researchers. *Development* **143**, 2677–2688 (2016).
25. Attanasio, C. et al. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006 (2013).
26. Tanaka, H. V. et al. A developmentally regulated switch from stem cells to dedifferentiation for limb muscle regeneration in newts. *Nat. Commun.* **7**, 11069 (2016).
27. Sánchez Alvarado, A. Developmental biology: a cellular view of regeneration. *Nature* **460**, 39–40 (2009).
28. Kragl, M. et al. Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* **460**, 60–65 (2009).
29. Purnell, B. A. Regrow like an axolotl. *Science* **355**, 592 (2017).
30. Nacu, E., Gromberg, E., Oliveira, C. R., Drechsel, D. & Tanaka, E. M. FGF8 and SHH substitute for anterior–posterior tissue interactions to induce limb regeneration. *Nature* **533**, 407–410 (2016).
31. Roensch, K., Tazaki, A., Chara, O. & Tanaka, E. M. Progressive specification rather than intercalation of segments during limb regeneration. *Science* **342**, 1375–1379 (2013).
32. Kaufman, C. K. et al. A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* **351**, aad2197 (2016).
33. Coufal, N. G. et al. L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
34. Ivancevic, A. M., Kortschak, R. D., Bertozzi, T. & Adelson, D. L. LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome Biol. Evol.* **8**, 3301–3322 (2016).
35. Bridier-Nahmias, A. et al. Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. *Science* **348**, 585–588 (2015).
36. Mashanov, V. S., Zueva, O. R. & García-Arrarás, J. E. Retrotransposons in animal regeneration: overlooked components of the regenerative machinery? *Mob. Genet. Elements* **2**, 244–247 (2012).
37. Baillie, J. K. et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534–537 (2011).

Acknowledgements We thank J. Wysocka for her review of the manuscript and helpful suggestions. We thank the Stanford Functional Genomics Facility, Stanford Cell Sciences Imaging Facility, Lorry Lokey Imaging Facility, and Stanford Shared FACS Facility Cores. We thank D. J. Hunter and D. Atashroo for their respective contributions to the design of the distraction device. This work was supported by the National Institutes of Health (NIH) grants R01DE026730 (to M.T.L. and R.C.R.), U24DE026914 (to M.T.L.) and K08DE024269 (to D.C.W.); the Child Health Research Institute (CHRI) at Stanford University (D.C.W.); The Hagey Laboratory for Pediatric Regenerative Medicine (M.T.L.); the Steinhart/Reed Award (M.T.L.); the Gunn–Oliver Fund (M.T.L.); and NIH grant P50-HG007735 and the Scleroderma Research Foundation (H.Y.C.). H.Y.C. is an Investigator of the Howard Hughes Medical Institute.

Reviewer information *Nature* thanks C. Tabin, L. Gerstenfeld and P. Scacheri for their contribution to the peer review of this work.

Author contributions R.C.R. conceived the study and performed microsurgical procedures. R.C.R. and A.C.C. produced figures, wrote the manuscript, and performed ATAC-seq and RNA-seq experiments. A.C.C. performed ATAC-seq and RNA-seq analysis. R.C.R. and A.Sa. performed FACS isolation and experiments based on skeletal stem cells and progenitor cells. Y.W. assisted with ATAC analysis of transposable elements. T.L., O.M., M.L.L., C.D.M., M.P.M., E.Z.S., R.E.J., A.Sh., C.K.F.C. and D.C.W. generated key materials and executed multiple experiments. O.D.K. provided key materials and reagents and support. H.Y.C. and M.T.L. oversaw the work and provided support. All authors reviewed the manuscript and discussed the work.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0650-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0650-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to H.Y.C. or M.T.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Animals. Rainbow mice³⁸ containing reporter gene $R26^{VT/GK3}$ were crossed with $Actin^{creERT2}$ and $Sox9^{creERT2}$ mice (Jackson Laboratories) to obtain $Actin^{creERT2};R26^{Rainbow}$ and $Sox9^{creERT2};R26^{Rainbow}$ mice. Additionally, mice containing $R26^{mT/mG}$ (Jackson Laboratories) were crossed with $Sox9^{creERT2}$ and $Sox10^{creERT2}$ animals to obtain $Sox9^{creERT2};R26^{mT/mG}$ and $Sox10^{creERT2};R26^{mT/mG}$ mice (Jackson Laboratories). Tamoxifen (Sigma-Aldrich) was dissolved in 90% corn oil/10% ethanol (vol./vol.), and filtered through a 0.2- μ m membrane. Ten-week-old male mice were given intraperitoneal injections of 200 mg per kg body weight of tamoxifen daily for five consecutive days. After one week, distraction osteogenesis was surgically applied to hemimandibles as specified. For all conditions, age- and sex-matched littermates served as control animals and were given corn-oil injections (no tamoxifen). Ten-week old male C57BL/6J mice (Jackson Laboratories) were used for isolation of mouse SSCs and skeletal progenitors for gene-expression and ATAC-sequencing experiments as indicated. Ten-week-old male C57BL/6J-Tg(CAG-EGFP)10sb/J mice (Jackson Laboratories) were used for parabiosis experiments.

All experiments were performed in accordance with Stanford University Animal Care and Use Committee guidelines. Animals were housed in a light- and temperature-controlled environment and given food and water ad libitum. Sample size was no less than three animals for all experiments. Specific numbers are stated in the figure legends. Sample size was determined by the number of viable animals of the right age and genotype at the time of experiment. All data analysis was conducted in a blinded manner for experiments in which the investigator could affect the outcome, such as any μ CT analyses, cell counting in immunofluorescence-based assays, assessment of molecular treatments, clonal analyses related to lineage tracing, and so on. Animals with the appropriate genotype were randomly allocated to experimental conditions.

Device design and manufacturing. Mandibular distraction devices were manufactured via computer-aided design (CAD) in SolidWorks (SolidWorks) for 3D printing (ProJet 3510 HD Plus, 3DSYSTEMS) at 16- μ m resolution.

Mandibular distraction surgery. Animals were divided into four groups in this study: sham-operated, fractured, acutely lengthened and gradually distracted. In brief, animals were anaesthetized (with 20 mg kg⁻¹ Ketaset, 1.5 mg kg⁻¹ xylazine and 0.2 mg kg⁻¹ acepromazine maleate), given a preoperative dose of antibiotics (10 mg kg⁻¹ cefazolin) and prepped with Betadine, and their incisors were clipped. An incision was made over the right hemimandible, the masseter muscle divided, and the mandible exposed. One 0.6-mm hole was drilled 3 mm anterior, and one 3 mm posterior, to a line dividing the mandibular ramus just posterior to the third molar. An osteotomy was then performed posterior to the third molar using a diamond disc saw under constant saline irrigation (Brasseler). Distraction plates were secured with insertion of tight-fit 0.65-mm screws (McMaster-Carr). The muscle and skin were then closed in layers. All animals tolerated the procedure well and received appropriate postoperative analgesia. Postoperative mortality was 5%, and all deaths were replaced with new animals to obtain the final numbers.

Mandibular distraction protocol. The gradual-distraction protocol consisted of a 5-day latency period after the initial osteotomy and fixation of the distraction device, followed by 10 days of distraction at a rate of 0.15 mm every 12 h (for a total of 3.0 mm) and then 28 days of consolidation. For our acute-lengthening protocol, lengthening was performed equal to the total distraction amount (3.0 mm) following a 5-day latency period, with a consolidation period ending at 43 days postoperation. All specimens were collected at either 29 days or 43 days postoperation, the latter being the end of bone consolidation.

Micro-computed tomography scanning. Devices were removed carefully before fixation. Specimens were manually palpated as a screening test for complete bone union before overnight fixation (in 2% paraformaldehyde (PFA) at 4°C), then were processed to 70% ethanol and scanned using a MicroXCT-200 (Carl Zeiss Microscopy) at 40 kV and 160 μ A. We took 1,200 projection images at a total integration time of 6 s, with linear magnification of $\times 2$ and a pixel size of 10 μ m. The morphological data were reconstructed in a 3D solid volume in the 3D image data visualization program Avizo Fire (FEI). Following μ CT scanning, tissues from POD43 were processed for histomorphometric and histologic evaluation. A total of six ten-week old C57BL/6J male mice were included per treatment group ($n = 6$). **Histological analysis.** Dissected specimens were fixed in 2% PFA at 4°C overnight. Specimens were decalcified in 400 mM EDTA in phosphate-buffered saline (PBS; pH 7.2) at 4°C for four weeks, with a change of EDTA every 48 h. The specimens were then dehydrated in 30% sucrose at 4°C overnight. Specimens were then embedded in optimal cutting temperature (OCT) compound and sectioned at 8 μ m. Representative sections were stained with Movat's modified pentachrome or with alizarin red depending on the individual experiment.

Histomorphometric analysis. Histomorphometry was performed using micrographs of Movat's pentachrome staining obtained for subsequent quantitative analysis of relative bone (yellow) and cartilage (blue) formation in the distracted callus. A region of interest (ROI) was assigned in Adobe Photoshop for pixel-based

quantification (bone pixels/total ROI pixels; cartilage pixels/total ROI pixels; stromal tissue pixels/total ROI pixels) using ten slides (one every 50 μ m) per specimen across five specimens.

Immunofluorescence. Immunofluorescence on cryopreserved ectopic bone specimens were performed using a mouse-on-mouse immunodetection kit from Vector Laboratories according to the manufacturer's instructions. In brief, specimens were treated with a blocking reagent, then probed with primary antibody at 4°C overnight. Specimens were next washed with PBS, probed with Alexa-dye-conjugated antibodies, washed, cover-slipped and imaged with a Leica DMI6000B inverted microscope system.

Tamoxifen induction of the Rainbow reporter system. We used Rainbow mice for clonal analysis of bone regeneration during mandibular distraction. Rainbow mice were crossed with the ubiquitous $Actin^{creERT2}$ or $Sox9^{creERT2}$ driver so as to mark all cells or skeletal-lineage-derived cells after systemic tamoxifen induction by intraperitoneal injection. The Rainbow reporter ($R26^{VT/GK3}$) is a multicolour Cre-dependent marker system with a four-colour reporter construct in the ROSA locus. Once recombination occurs, cells are randomly and genetically marked with one of ten possible colour combinations and daughter cells will be marked with the same colour as the parent cell, creating a fluorescent mosaic pattern upon analysis. Nine-week old male $Actin^{creERT2};R26^{Rainbow}$ or $Sox9^{creERT2};R26^{Rainbow}$ mice were administered intraperitoneal injections with tamoxifen at 200 mg kg⁻¹ daily for five consecutive days. At ten weeks of age, distraction osteogenesis was surgically applied to hemimandibles in systemically induced Rainbow mice. Hemimandibles were collected 15 and 29 days after initial surgery and immediately fixed in 2% paraformaldehyde (wt/vol.) overnight at 4°C in the dark. Tissue samples were prepared for cryo-embedding by soaking in 30% (vol./vol.) sucrose in PBS at 4°C for 24 h. Samples were removed from the sucrose solution and tissue blocks prepared by embedding in Tissue Tek OCT (Fisher). Cryosections were obtained and counterstained for nuclei with Hoechst 33342 dye for subsequent confocal microscopy (Leica TCS Sp8).

Imaging analysis. Laser scanning confocal microscopy was performed with a Leica TCS SP8 X confocal microscope (Leica Microsystems) with an objective lens ($\times 10$ HC PL APO, air, numerical aperture 0.40; $\times 20$ HC PL APO IMM CORR CS2, H₂O/glycerol/oil, numerical aperture 0.75), located in the Cell Sciences Imaging Facility (Stanford University). Raw image stacks were imported into Fiji (NIH) or Imaris (Bitplane/Perkin Elmer) software for further analysis. Imaris software was used to analyse cells and obtain xyz coordinates of individual clones. All clones were examined individually to confirm that they reported a single colour. Only clones that could be visually determined to consist of five cells or more were included in the analysis of clone size. The theoretical recombination frequency resulting in a cell being marked by one colour in Rainbow mice is 1/10 possible combinations. The probability of obtaining by chance single-colour clones containing five or more cells would then be approximately greater than or equal to 1 in 2.1 million. Therefore, we included only clones of five cells or more in our analysis. Imaris (Bitplane) software was used to render 3D volumes of z-stacks and to perform image analysis. Cell-spotting analysis of Rainbow fluorescent clones was performed using the volume surface and spot creation tools, thresholding by volume and signal quality. For Rainbow clonal analysis, adjacency was determined by thresholding maximum distance to the nearest cell using the spot-to-spot distance tool. Quantification of images was performed on Imaris using the spots rendering statistics tool as per the manufacturer's protocol, within a 1–3-mm-wide ROI containing the distraction gap and callus. For the osteotomy site, a surface was manually defined in Imaris by the edge of dissection within the callus area. The Imaris distance transformation module from their XTensions library was then used to determine the vector of each clone for a determination of angles, and the axis of each clone was measured with respect to the osteotomy edge. For determination of angular expansion, clones of at least ten cells were evaluated throughout five distracted mandibles for each group. Measured clonal vectors were plotted within each 10° segment. Provided images are typically presented as a maximal projection of either 8–12- μ m optical sections or 30–50- μ m whole-mount renderings, unless otherwise specified. For visualization of individual labelled cells expressing the Rainbow reporter, the brightness and contrast were adjusted accordingly for the green (eGFP), blue (mCerulean), orange (mOrange) and red (mCherry) channels, and composite serial image sequences were assembled. Tiled images were stitched by a grid and collection stitching plugin in Fiji.

Tissue-specific Rainbow clonal analysis. To further explore the periosteum as a cellular source of progenitors during mandibular distraction, we devised a strategy of local genetic labelling of the mandibular periosteum for tissue-specific clonal analysis. To induce site-specific recombination in $Actin^{creERT2};R26^{Rainbow}$ mice, we delivered activated 4-hydroxytamoxifen (10 μ g) in a liposomal formulation (20 μ g μ l⁻¹) to the periosteum on the buccal surface of the mandible at a 1-mm distance posterior to the third molar. After one week, tissues were collected to evaluate the local labelling strategy in intact mandibles, at which time distraction osteogenesis was surgically applied to hemimandibles in locally induced Rainbow

mice and tissues were harvested at POD15. Periosteum-specific clonal analysis was performed with four animals (including controls receiving liposomal formulation without 4-hydroxytamoxifen) across two independent experiments ($n = 8$) using *Actin^{creERT2};R26^{Rainbow}* mice.

Whole-mount Rainbow clonal analysis. Nine-week-old male *Actin^{creERT2};R26^{Rainbow}*, *Sox9^{creERT2};R26^{Rainbow}* and *Sox9^{creERT2};R26^{Rainbow};Ptk2^{fl/fl}* mice were intraperitoneally induced with tamoxifen as described. Intact periosteum (one-year lineage trace) or distraction calluses (PODs 5, 8, 10, 12 and 15) were collected using fine-precision surgical techniques to expose the mandibular periosteum. Upon exposure, mandibular periosteum was preserved in Fluoromount-G medium (SouthernBiotech) and mounted onto microscope slides (Fisher Scientific). Preserved periosteum specimens were imaged immediately using confocal microscopy techniques as described above. Whole-mount Rainbow clonal analysis was performed using a minimum of ten biological replicates for *Actin^{creERT2};R26^{Rainbow}*, and ten biological replicates for each of *Sox9^{creERT2};R26^{Rainbow}* and *Sox9^{creERT2};R26^{Rainbow};Ptk2^{fl/fl}* animals.

Lineage tracing of Sox-expressing progenitors in vivo. To evaluate Sox-factor-expressing cells in mandibular distraction, we used *Sox10^{creERT2};R26^{mT/mG}* mice to trace progenitors. Nine-week-old male *Sox10^{creERT2};R26^{mT/mG}* mice were administered intraperitoneal injections with tamoxifen as described above. The hemimandibles of *Sox10^{creERT2};R26^{mT/mG}* mice were isolated and transverse sections prepared for confocal microscopy as described above. Lineage tracing of Sox-expressing progenitors was performed using a minimum of six separate animals as biological replicates ($n = 6$).

Parabiosis for circulating cell fate in distraction. To determine the contribution of circulating cells to newly formed bone, we carried out parabiosis as previously described³⁹. In brief, sex- and age-matched GFP-labelled mice (C57BL/6 *J-Tg(CAG-EGFP)10sb/J*; Jackson Laboratories) and non-GFP littermates (C57BL/6 wild-type; Jackson Laboratories) were used. Animals were anaesthetized via inhalational anaesthesia. An incision in the skin was made from the base of the right foreleg to the base of the right hind leg of one parabiont, and from the left foreleg to the base of the left hind leg of the other parabiont. The skins were sutured together at the foreleg and hind-leg joints. The remaining dorsal and ventral flaps were stapled together. Analgesia was administered postoperatively. After one month of parabiosis, peripheral samples were collected from the tail, and FACS was used to assess parabiont blood chimaerism. After peripheral blood chimaerism reached a 1/1 ratio—representing a complete fusion of the circulatory systems of both parabionts—distraction osteogenesis was performed on non-GFP mice as described above. Hemimandibles were collected to analyse the distracted callus for the presence of GFP cells after two weeks of consolidation (POD29), when bone union was complete. Tissue processing and histology was performed on hemimandibles as previously described. Parabiosis experiments were performed using four biological replicates ($n = 4$) with data represented as means \pm s.d.

Sample preparation and FACS isolation. Bones were dissected and serially digested in collagenase digestion buffer supplemented with DNase at 37 °C for 40 min under constant agitation; total dissociated cells were filtered through 40- μ m nylon mesh, pelleted at 200g at 4 °C, resuspended in staining medium (2% fetal calf serum (FCS) in PBS), and stained with fluorochrome-conjugated antibodies against CD45, Ter119, CD202b, Thy1.1, Thy 1.2, CD105, CD51 and 6C3, and with a streptavidin-conjugated antibody for CD200. Propidium iodide staining was performed to exclude dead cells. FACS analysis was performed on a FACS Aria II Instrument (BD Biosciences) using a 70- μ m nozzle. Gating schemes were established with fluorescence-minus-one controls and propidium iodide was used for viability staining. All cell populations were double-sorted for purification and subsequently evaluated for their functional responses as outlined below. Flow-cytometry plots are representative of a minimum of three independent experiments.

To calculate cell population frequencies (Figs. 3e, 5l), we assessed five post-operative calluses ($n = 5$) by FACS and represent data as an average across three independent experiments. Single uninjured mandibles (dissected anterior of condyle but posterior to third molar) contain approximately 1×10^3 to 1.2×10^3 SSCs, 2.8×10^3 to 3.5×10^3 BCSPs, and 4.5×10^3 to 5×10^3 osteoprogenitors. All molecular and flow-cytometric analyses of FACS-purified cell populations in the SSC hierarchy throughout this study were performed using double-sorted cells to ensure purity of each population. The double-sorting technique reduces the yield of purified populations by roughly one half, necessitating surgical operations on approximately twice the number of animals expected to achieve the desired cell number for a given experiment (see the Methods sections below on ATAC-seq and RNA-seq). To double-sort SSCs and BCSPs, we first performed FACS isolation of the 'double-negative' population (CD45⁻ Ter119⁻ CD202b⁻ Thy1⁻ 6C3⁻ CD51⁺) on the basis of yield, and then re-sorted on the basis of purity for SSCs (CD45⁻ Ter119⁻ CD202b⁻ Thy1⁻ 6C3⁻ CD51⁺ CD105⁻ CD200⁺) and BCSPs (CD45⁻ Ter119⁻ CD202b⁻ Thy1⁻ 6C3⁻ CD51⁺ CD105⁺). For double-sorting of terminally differentiated Thy1⁺ populations, we first performed FACS isolation of the 'Thy'

population (CD45⁻ Ter119⁻ CD202b⁻ Thy1⁺ 6C3⁻ CD51⁺) on the basis of yield, and then re-sorted cells on the basis of purity for osteoprogenitors (CD45⁻ Ter119⁻ CD202b⁻ Thy1⁺ 6C3⁻ CD51⁺ CD105⁻ CD200⁻) and chondrocytic progenitors (CD45⁻ Ter119⁻ CD202b⁻ Thy1⁺ 6C3⁻ CD51⁺ CD105⁺ CD200⁺).

EdU analysis in vivo. Mice were injected intraperitoneally with 100 mg kg⁻¹ EdU (Life Technologies) 12 h before euthanasia. Cells were double-sorted for purity through FACS isolation as described above before being fixed and subject to EdU staining using the Click-iT Plus EdU Alexa-488 flow-cytometry assay kit (Life Technologies). Propidium iodide was used to stain for total DNA viability content and the percentage of GFP-negative and GFP-positive cells showing EdU incorporation was analysed using a BD FACSaria II. SSCs were isolated from ten-week-old male mice according to strain, and colony-formation assays were performed in triplicate, using a minimum of three biological replicates across three independent experiments. Graphs depict means \pm s.d. across the biological replicates.

Colony formation in vitro. Isolated SSCs from uninjured wild-type or FAK(cKO) (uninduced, no tamoxifen) mice were directly plated onto precoated (0.1% gelatin) culture plates (100 cells per well in a 10 cm² well plate) in MEMa medium with 20% FCS under low O₂ (2% atmospheric oxygen, 7.5% CO₂) conditions. After four days of attachment, SSCs from wild-type mandibles were treated with the FAK inhibitor PF-573228 (FAKi) or with DMSO (control), while SSCs from the FAK(cKO) mice were treated with activated tamoxifen (4-OHT) or control medium every other day for one week. SSCs isolated from uninjured wild-type mandibles for serial colony-forming units, and plated in the same conditions described above, were not treated with any substance. Colony-forming units were identified using an inverted microscope under $\times 40$ magnification. Specimens were examined under phase microscopy and a cloning ring was used for quantification. Colonies were assessed for size and cell morphology as previously described⁶. The cells were subsequently lifted for staining and analysis by FACS, or plated for tertiary colonies. Similarly, isolated SSCs from uninjured wild-type mice were plated onto precoated dishes (1 cell per well in a 4-cm² well plate), cultured and analysed in identical conditions to those described above. SSCs were isolated from ten-week-old male mice according to strain, and colony-formation assays were performed in triplicate using a minimum of three biological replicates across three independent experiments ($n = 9$); graphs depict means \pm s.d. across the biological replicates.

Osteogenic differentiation assay. Upon isolation of cells by FACS, colonies were grown over two weeks as described above. Each cell-type condition (f-SSC and d-SSC) was incubated with osteogenic medium for two weeks with the medium changed every other day. After undergoing two weeks of osteogenic differentiation, cells were washed with PBS followed by ultrapure water. The monolayer of cells was fixed with 100% ethanol for 15 min and stained with alizarin red solution for 1 h at room temperature. The cells were washed several times with ultrapure water and imaged for osteogenic potential immediately under a bright-field microscope. Following imaging, the cells were treated with a methanol/acetic acid mixture for 15 min and absorbance was detected with an Ultraspec 2100 UV/Visible Spectrophotometer (Biochrom, Harvard Bioscience) at 450 nm to measure alizarin red protein concentration across each cell type.

ATAC-seq. ATAC-seq was performed as previously described, using the Omni-ATAC protocol⁴⁰. In brief, 10,000 cells per replicate were pelleted and lysed in 50 μ l lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂) with 0.1% NP40, 0.1% Tween-20 and 0.01% digitonin for 3 min on ice. We then added 1 ml of cold lysis buffer plus 0.1% Tween-20, and centrifuged cells at 500g for 10 min at 4 °C. Following centrifugation, pelleted nuclei were resuspended in 50 μ l transposition mix (25 μ l 2 \times tagmentation DNA (TD) buffer, 2.5 μ l transposase (final concentration 100 nM), 16.5 μ l PBS, 0.5 μ l 1% digitonin, 0.5 μ l 10% Tween-20, 5 μ l H₂O) and incubated for 30 min at 37 °C with shaking at 1,000 r.p.m. Transposition mix was cleaned up using Qiagen MinElute columns. Library preparation was performed exactly as previously described¹³. To achieve adequate cell numbers (minimum 10,000 double-sorted cells per replicate), we required 18–22 total 'distracted' mandibles (one per animal) per replicate in a single experiment, and 26–30 total 'fracture' mandibles per replicate in a single experiment. Comparisons between surgical conditions were strictly performed using littermates randomly assigned and mandibles operated upon at the same date. Cells were FACS-isolated within the same session for parallel library preparation (for example, f- and d-SSC, BCSP and OP populations isolated on the same date for a direct comparison). A minimum of two biological replicates (as two independent operations in separate animal cohorts and separate FACS isolation experiments) were used for each condition in our ATAC-seq analyses.

ATAC-seq analysis. ATAC-seq libraries were sequenced on an Illumina NextSeq. Adaptor sequences were trimmed using CutAdapt software and reads were then mapped to the genome (mm9 build) using Bowtie2. Mapped reads were filtered for quality, and then reads mapping to segmental duplications and non-unique sequences were removed. Duplicates were further filtered out using PICARD. Peaks were called on each individual sample using MACS2 with a q -value cut-off of 0.01

and with no shifting model. Peaks from all samples were merged and bedtools was used to calculate the read depth at each peak for each sample. DESeq2 was used for pairwise differential peak calling with a cut-off of fold change >2 and $\text{padj} < 0.05$ for differential sites. To calculate P values in DESeq2, for each gene the counts were modelled using a generalized linear model (GLM) of negative binomial distribution among samples. The two-sided Wald statistics test was processed for significance of the GLM coefficients. The Benjamini–Hochberg correction was applied to all P values to account for the multiple tests performed. This gives the final adjusted P values used for assessing significance. For motif-enrichment analysis, we used HOMER v4.8.3 with default settings. The de novo motif search results are shown. For all motif enrichment, background sets of all peaks were used. For genomic annotations, HOMER's Annotate peaks tool was used.

Analysis of GO terms from ATAC-seq data. We carried out GO-term analysis of ATAC-seq sites using GREAT v3.0.0 (ref.⁴¹; <http://great.stanford.edu/public/html/>), with a background set of all peaks for a given cell type.

Analysis of retroviral elements in ATAC-seq data. To analyse retroviral-element accessibility in ATAC-seq data, we mapped all ATAC-seq reads to a bowtie2 index containing the sequences of all retroviral insertions from the RepeatMasker database (<http://www.repeatmasker.org/>). Reads mapping to each retroviral-element family were summed for each sample. Transfer RNAs and simple repeats were removed. The read counts were then normalized for each repeat type using DESeq2.

ATAC-seq data from neural crest cells. ATAC-seq data from NCC populations were downloaded from the Gene Expression Omnibus (accession number GSE89436). Data were downloaded in raw form (fastq) and processed using the same pipeline as above.

Small-molecule inhibition of FAK. Mice undergoing distraction were treated with local subcutaneous injections of the FAK inhibitor PF-573228 (Tocris) at a concentration of 50 μM . PF-573228 specifically inhibits the phosphorylation of FAK at tyrosine 397—the active site required for canonical activation of the FAK pathway in the context of integrin-based signalling events. We determined the minimum effective dose as the minimum required to inhibit FAK phosphorylation in vitro and in vivo, in both SSCs and BCSPs, without disrupting cell viability according to increased propidium iodide staining via FACS. We evaluated both SSCs and BCSPs after in vitro treatment using FAKi at 10 μM , 20 μM , 40 μM , 50 μM , 70 μM , 80 μM or 100 μM , compared with DMSO treatment alone. FAKi was added at each 24-h medium change over the course of one week. Using intracellular FACS and immunocytochemical staining for phosphorylated FAK, we determined that the 50 μM treatment substantially inhibited phosphorylation without disrupting cell viability when compared with DMSO treatment alone. Ex vivo intracellular FACS and immunocytochemical staining confirmed inhibition of FAK phosphorylation (Extended Data Fig. 9e) without disrupting cell viability when FAKi was administered subcutaneously every 24 h at a concentration of 50 μM in a 100- μl volume in vivo. We determined the minimum toxic concentration to be 70 μM in vitro and 80 μM in vivo, according to propidium iodide staining on FACS. To determine the effect of inhibiting the mechanotransduction pathway on the rate of bone healing during distraction osteogenesis, we administered wild-type C57BL/6 mice with daily local injections (100 μl total per day) of the FAK small-molecule inhibitor or control vector (DMSO) throughout the distraction period.

Conditional knockout of FAK. To determine whether FAK-pathway inhibition affects the functional response of skeletal stem and progenitor cells during mandibular distraction, we administered FAK(cKO) mice (*Sox9^{creERT2};Ptk2^{fl/fl}*) with injections of either tamoxifen (200 mg kg^{-1} in corn oil) or corn oil for five consecutive days. Distraction osteogenesis was then applied to hemimandibles of FAK(cKO) (tamoxifen) and control (no tamoxifen) littermates.

RNA-seq library preparation. RNA was extracted from FACS-sorted SSCs, BCSPs and osteoprogenitors using the Qiagen RNeasy Micro kit. For library preparation we modified the SMART-seq protocol as follows: following extraction, 1.5 ng of total RNA was reverse transcribed using an oligo-dT primer to make complementary DNA from polyadenylated transcripts. Reverse-transcribed cDNA was then pre-amplified for 11 PCR cycles. Next, 1 ng of cDNA was transposed with 5 μl of Nextera amplicon tagment mix for 5 min at 55 °C. Transposed libraries were then amplified for 11 cycles using Nextera polymerase chain reaction (PCR) master mix and adapters. Libraries were sequenced (single-end 75-base-pair reads) on an Illumina NextSeq 500 instrument.

RNA-seq analysis. Reads in fastq format were mapped and assigned to a transcript GTF file from UCSC and RefSeq using Tophat 2 (version 2.1.1). Reads in each transcript were then counted using FeatureCounts (featureCounts.v4.R) in R. For global clustering and PCA, read counts per transcript were normalized using DESeq2 and plotted using heatmap.2 and plotPCA(DESeq2) all in R. For differential-expression analysis, DESeq2 was used to make pairwise comparisons using default normalization parameters. Significantly differentially expressed genes are those showing a fold change of 1.5 or greater and an adjusted P value of less than 0.05. David V.6.8 was used for GO-term analysis. To achieve adequate cell numbers (minimum 5,000 cells per replicate), 18 total distracted mandibles (one per animal) were required to generate two technical replicates at each time point. Comparisons between postoperative time points were strictly performed using littermates randomly assigned and mandibles operated at successive time points leading up to the same collection or FACS isolation date. Cells were FACS-isolated within the same session for parallel library preparation. A minimum of two biological replicates (as two independent operations in separate animal cohorts and separate FACS-isolation experiments) were used for each condition in our RNA-seq analyses.

Immunofluorescence staining in vitro. SSCs were sorted through FACS isolation and culture for colonies from two to four weeks as described above. Cells were washed with PBS and fixed in 2% PFA then permeabilized with Triton X 100. Specimens were treated with a blocking reagent, then probed with primary antibody at 4 °C overnight. Specimens were washed with PBS, probed with Alexa-dye-conjugated antibodies, and again washed in PBS. Cells were stained with Hoechst dye and mounted on cover slides to be imaged with a Leica DMI6000B inverted microscope system.

Antibodies used were as follows: rabbit anti-mouse-FAK, Abcam ab81298, clone EP2160Y, lot GR237911-21; rabbit anti-mouse-S100A4, Abcam ab41532, lot GR322940-2; rabbit anti-mouse myelin proteolipid protein, Abcam ab28486, lot GR268116-9; rabbit anti-mouse-CDC42, Abcam ab155940, lot GR3177223-5; rabbit anti-mouse-Twist2, Biologicals, Lot R87976; rabbit anti-mouse-ZEB2, ThermoFisher PA5-20980, lot QL2127782; goat anti-rabbit immunoglobulin G with conjugated Alexa Fluor 488, Thermo Fisher A-11034, lot 1885241.

Statistical analysis. Data are expressed either as absolute numbers or as percentages \pm s.d. Statistical significance was assigned for $P \leq 0.05$. Data analysis was performed using Student's t -test assuming two-tailed distribution, and/or one-way analysis of variance (ANOVA) and post hoc Tukey correction. Note that $*P < 0.05$ to $****P < 0.0001$ indicate a significant difference. Statistical calculations were performed using the Prism software package (GraphPad), except where stated otherwise. No statistical method was used to predetermine sample size. Figure panels showing representative images are representative of at least two independent experiments and up to five, as indicated in the figure legends. Flow-cytometry plots are representative of at least eight independent experiments and up to twelve, as indicated in the figure legends.

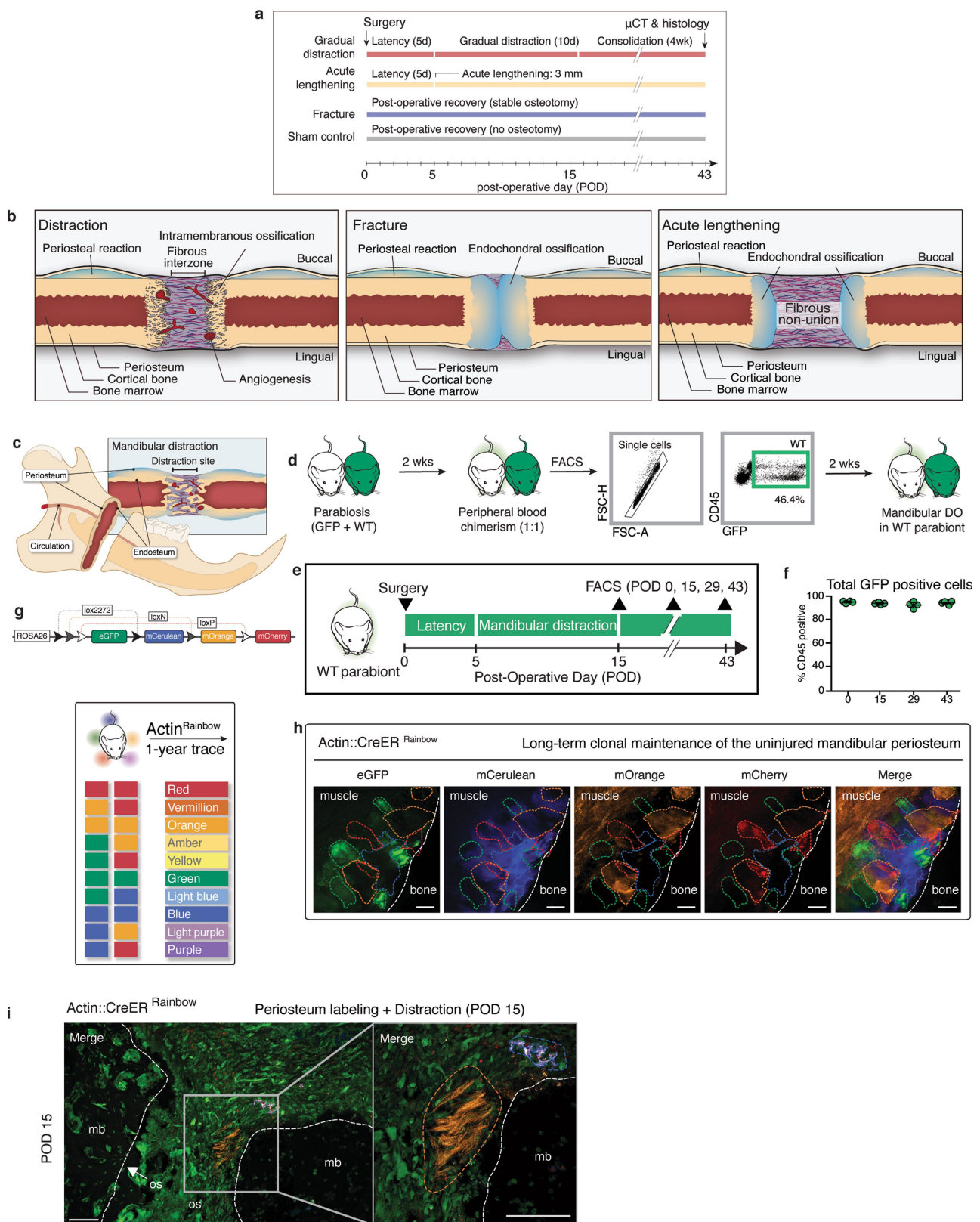
Statistics and reproducibility. For all figures, n indicates the number of animals per independent experiment unless otherwise indicated. All experiments were performed three times unless otherwise indicated. For all graphs, values plotted are the means with errors bars representing \pm the s.d.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data to support the conclusions in this manuscript can be found in the figures. All source data for graphs are available in the online version of the paper. Any other data can be requested from the corresponding authors. All ATAC-seq and RNA-seq data can be accessed from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) with accession number GSE104473.

38. Ueno, H. & Weissman, I. L. Clonal analysis of mouse development reveals a polyclonal origin for yolk sac blood islands. *Dev. Cell* **11**, 519–533 (2006).
39. Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L. & Weissman, I. L. Physiological migration of hematopoietic stem and progenitor cells. *Science* **294**, 1933–1936 (2001).
40. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
41. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).



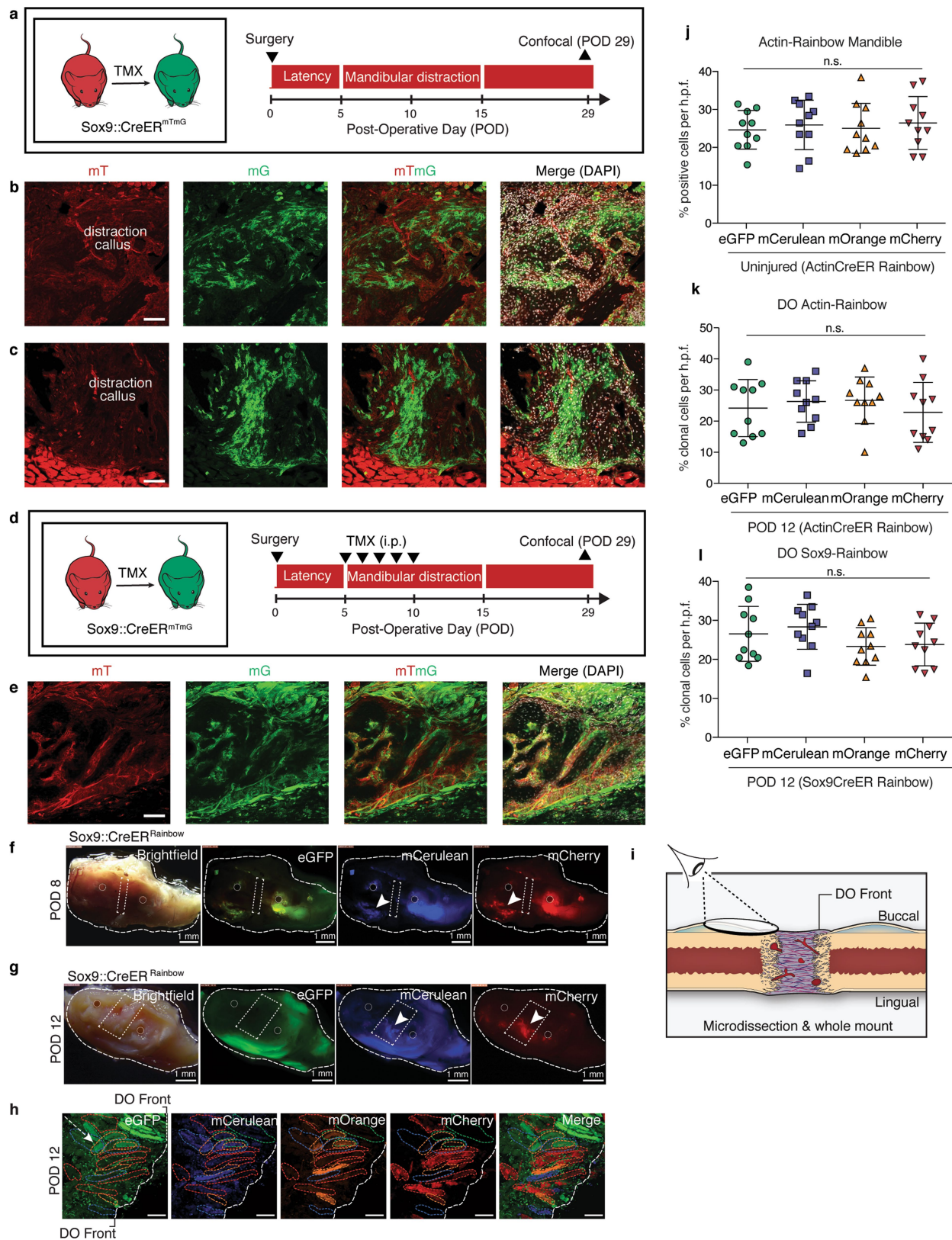
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Analysis of tissue sources of regeneration in distraction. **a**, Experimental timeline of distraction model.

b, Illustration of the tissue response (at POD15) in the mouse model of mandibular distraction (left), fracture (middle) and acute lengthening (right). **c**, Putative cellular sources of bone regeneration in mandibular distraction, including periosteum, endosteum and circulating progenitors.

d, Experimental scheme for detecting circulating progenitor cells in mandibular distraction. GFP mice are surgically fused to their wild-type (WT) littermates through parabiosis. Peripheral blood chimaerism is confirmed via flow cytometry. After GFP-positive cells are confirmed in the WT through FACS, mandibular distraction is performed on the WT parabiont. FSC-A, forward scatter area; FSC-H, forward scatter height. **e**, Upon detection of 1/1 blood chimaerism, WT parabionts undergo mandibular distraction according to the timeline outlined. **f**, Quantification of the haematopoietic fraction of GFP-positive cells

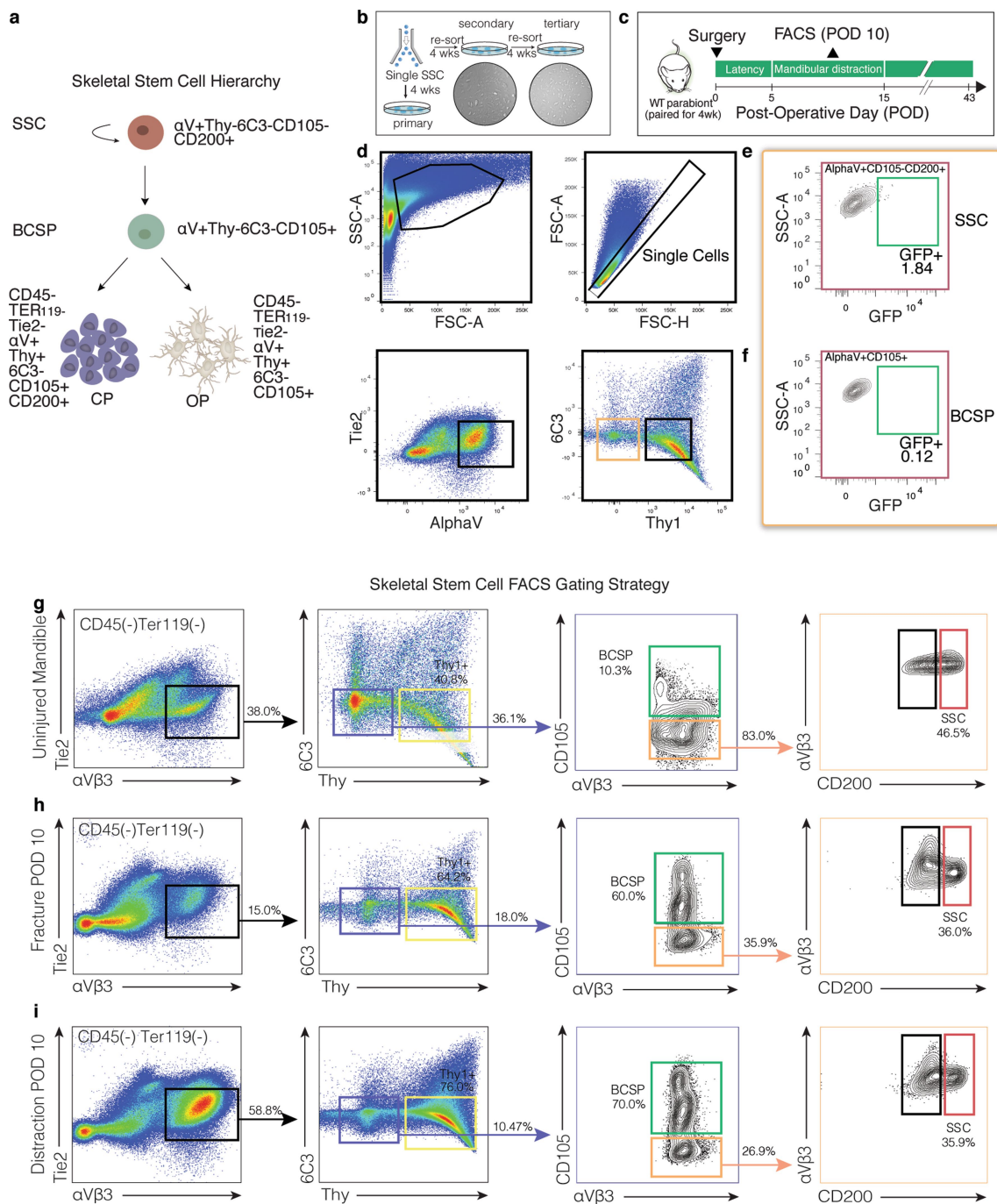
obtained from distraction calluses ($n = 4$ biological replicates per time point). **g**, Representation of the Rainbow reporter construct at the *R26* locus and the colours produced by random recombination. **h**, One-year tracing of mandibles under normal homeostasis (uninjured), with confocal micrographs of whole-mount periosteum one year after recombination. Clones are shown with coloured dotted outlines. The white dotted line demarcates skeletal muscle (upper left quadrant) from the periosteum. The view is a buccal-to-lingual view of the posterior periosteum overlying the body of the mandible. $n = 50$ clones, with 16–151 cells per clone. **i**, Confocal micrograph of the transverse mandible section from a Rainbow mouse at POD15 after targeted labelling of the periosteum for subsequent mandibular distraction ($n = 8$). Coloured outlines indicate single clones; the white dotted outline indicates mandibular bone (mb) at the distraction site. n refers to the number of animals in each independent experiment.



Extended Data Fig. 2 | See next page for caption.

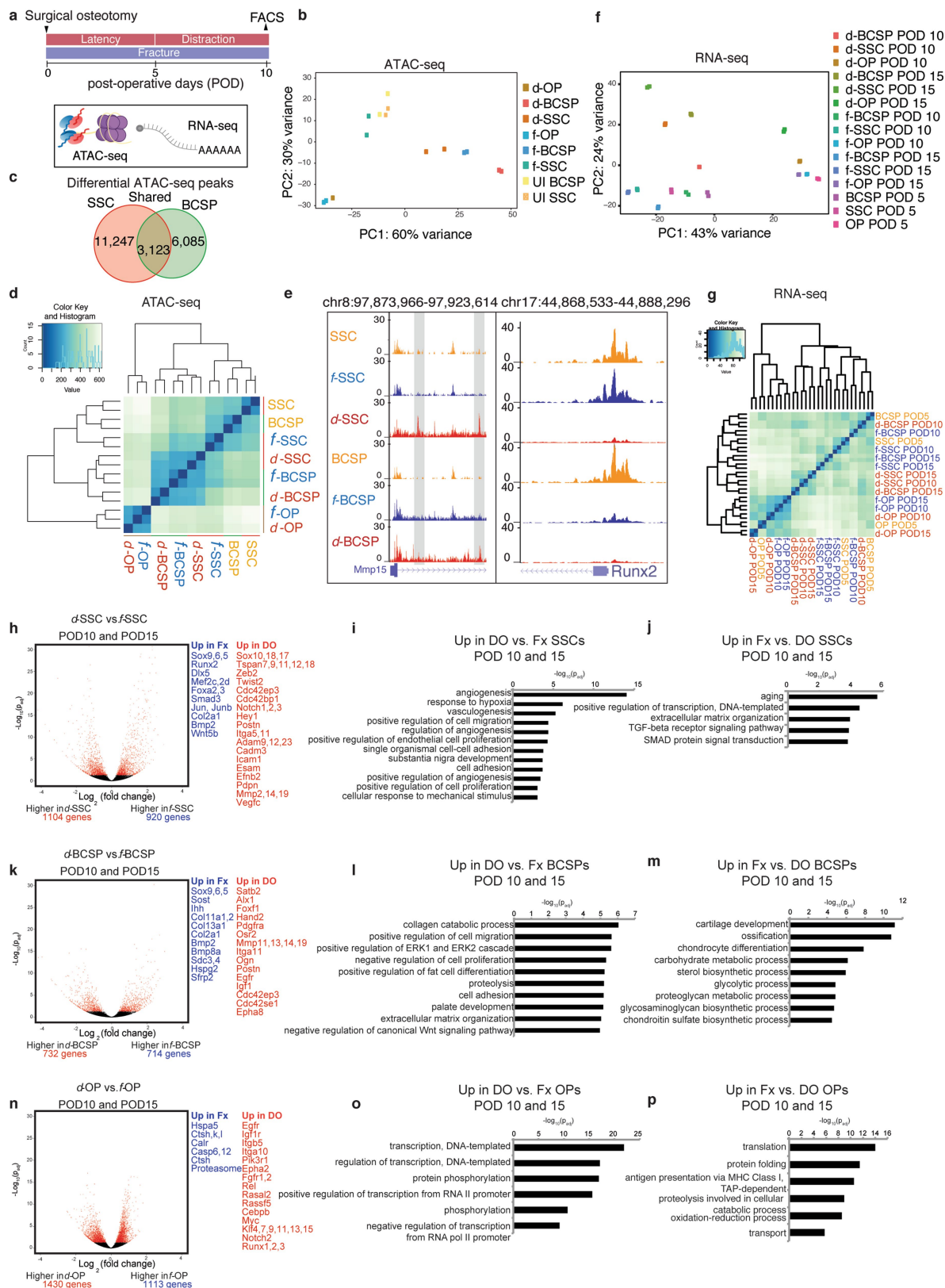
Extended Data Fig. 2 | Skeletal lineage tracing demonstrates labelling of bone regenerate in mandibular distraction. **a**, Experimental scheme for using tamoxifen (TMX; administered at 9 weeks of age, before surgery at 10 weeks) to induce recombination in *Sox9^{creERT2};R26^{mT/mG}* mice to carry out skeletal lineage tracing, with isolation of mandibles at POD29 ($n = 6$). **b**, Confocal micrograph (lingual mandible) of the distraction callus in *Sox9^{creERT2};R26^{mT/mG}* mice after TMX induction. Filters are shown in the following order from left to right: mT (mTomato, background), mG (mGFP, Sox9 lineage), mTmG (merged) and merged (mTmG with DAPI) ($n = 6$). **c**, As for **b**, but for the buccal mandible. **d**, Experimental scheme for tamoxifen induction of *Sox9^{creERT2};R26^{mT/mG}* mice during the early phase of distraction (POD5–10) for lineage tracing of Sox9⁺ cells during distraction. Mandibles were isolated at POD29 for confocal microscopy ($n = 4$). i.p., intraperitoneal. **e**, Confocal micrographs showing the contribution of the skeletal lineage (mGFP, Sox9 lineage) to new bone formed in mandibular distraction. **f**, Low-magnification

images of the *Sox9^{creERT2};R26^{Rainbow}* mandible during distraction at POD8, including bright-field, eGFP, mCerulean and mCherry filters under a dissection microscope. The induced Rainbow-coloured clones (white arrowheads) are anterior and posterior to the site of distraction (dotted white box) between the pin holes (dotted white circles) ($n = 8$). **g**, As for **f**, but for POD12. **h**, As for Fig. 11, but for POD12 ($n = 5$). **i**, Schematic showing the vantage point for clonal analysis of the skeletal-lineage (*Sox9^{creERT2};R26^{Rainbow}*) contribution to regeneration during distraction, through the use of whole-mount specimens from mandibular callus microdissection. DO, distraction osteogenesis. **j–l**, Quantification of clones (whole-mount imaging) that are positive for each fluorophore per high-pass filter (h.p.f.; $\times 40$ magnification) two weeks after tamoxifen induction (one injection per day, five days total) plus distraction until POD12 using *Actin^{creERT2};R26^{Rainbow}* mice (**j, k**) and *Sox9^{creERT2};R26^{Rainbow}* mice (**l**). Means \pm s.d. are shown Scale bars, 200 μ m (**b, c, e**), 1 mm (**f**). n refers to the number of animals in each independent experiment.



Extended Data Fig. 3 | Dynamics of SSCs and progenitor cells in parabiosis and distraction. **a**, The skeletal stem cell lineage, showing the immunological phenotype of each cell. CP, chondroprogenitor; OP, osteoprogenitor. **b**, FACS isolation of single SSCs for evaluation of serial colony-forming potential in vitro. Secondary (left) and tertiary (right) colonies are shown. **c**, Experimental scheme for detecting circulating cells in mandibular distraction at POD10. **d**, FACS gating of $Thy1^{-}6C3^{-}$ cells (light brown box) for detection of circulating SSCs (green box in **e**) and progenitor cells (green box in **f**) in mandibular distraction.

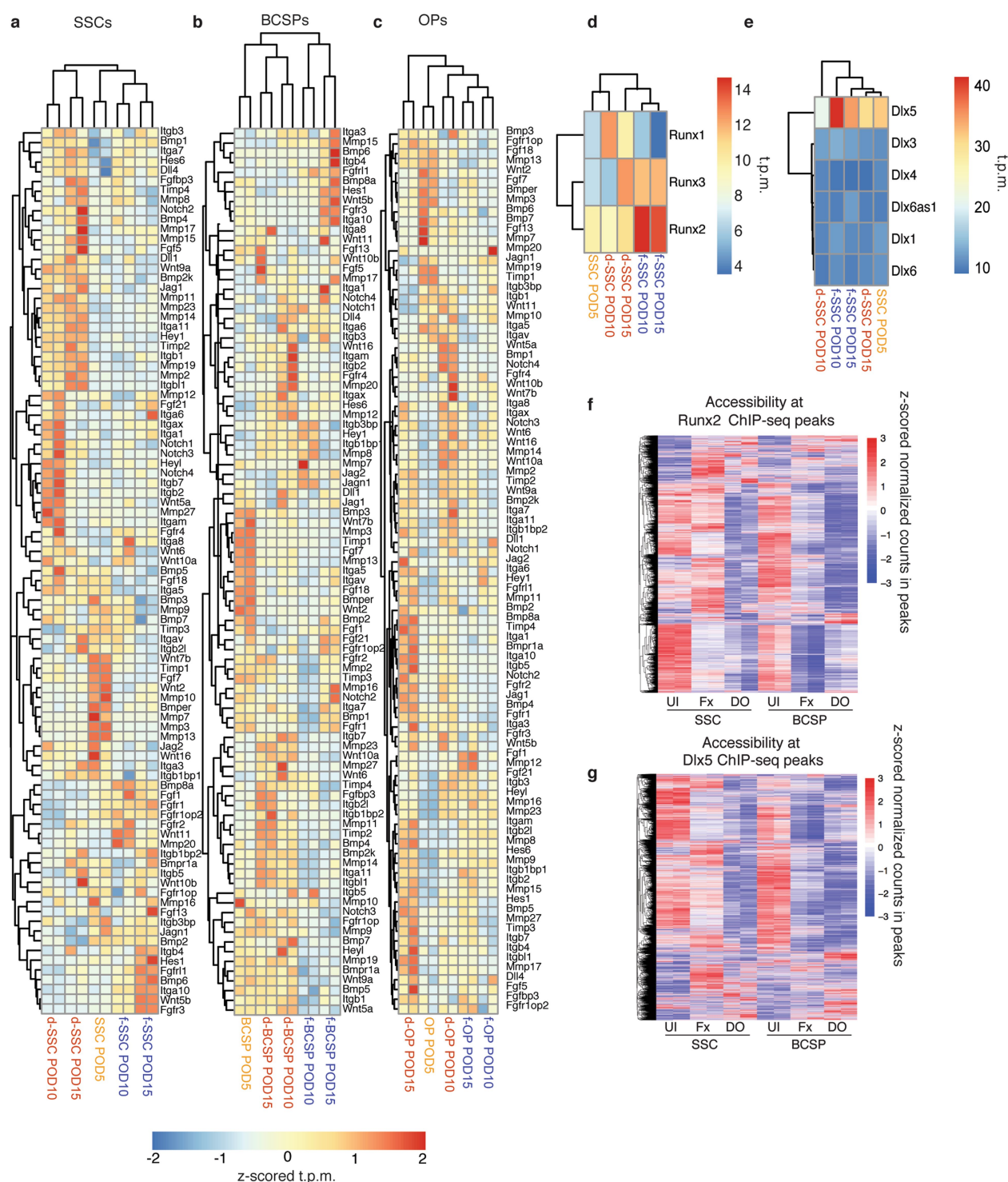
e, FACS gating shows an absence of circulating (GFP^{+}) SSCs in mandibular distraction ($n=8$; mean = 2.02 ± 0.78). **f**, As for **e**, but for BCSPs ($n=8$; mean = 1.83 ± 0.91). **g-i**, FACS isolation from uninjured (**g**; 38.0% $\alpha V\beta 3$ -positive events within the gate) versus fracture (**h**; 15.0%) and distraction (**i**; 58.8%) conditions reveal expansion of the SSC hierarchy (left column) in response to distraction by POD10. Representative of three independent experiments. n refers to the number of animals in each independent experiment.



Extended Data Fig. 4 | See next page for caption.

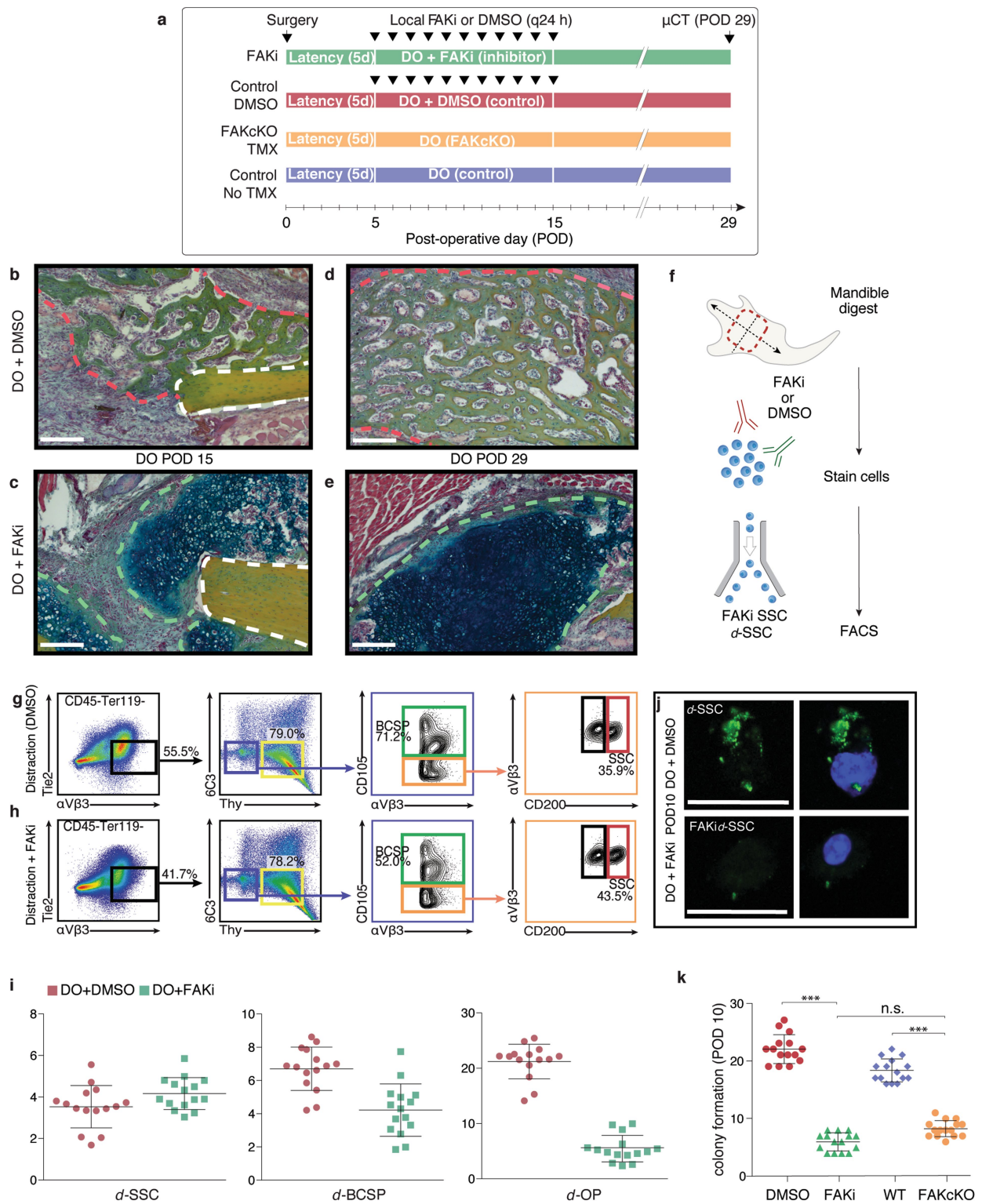
Extended Data Fig. 4 | Changes in gene regulation in response to distraction. **a**, Schematic showing FACS isolation of stem and progenitor cells from uninjured, fractured and distracted mandibles for ATAC-seq (assay for accessibility of chromatin, which is shown with purple spheres and yellow lines, to depict nucleosomes) and RNA-seq (shown with a polyA strand). Samples were collected in duplicate for ATAC-seq at POD10 and in duplicate for RNA-seq at PODs 5, 10 and 15. **b**, PCA showing PC1 and PC2 for all ATAC-seq data ($n = 164,266$ peaks). UI, uninjured. **c**, Venn diagram showing the number of differential peaks between the fracture and distraction conditions in both SSCs and BCSPs. Overlap between these (shared peaks; twofold change with $P < 0.05$ in both SSCs and BCSPs) is shown in the centre. **d**, Cluster dendrogram for ATAC-seq, showing clustering of samples based on all ATAC-seq peaks from SSCs, BCSPs and OPs. **e**, Example loci (*Mmp15* and *Runx2*, with their locations on chromosomes 8 and 17, respectively, shown at the top), revealing accessibility that is distraction-specific (*Mmp15*) and fracture-specific (*Runx2*), respectively. The height of the genome browser tracks shows the number of reads normalized by read depth and overall peak

enrichment in the library. **f**, PCA showing PC1 and PC2 for all RNA-seq data at PODs 5, 10 and 15 ($n = 17,491$ genes). **g**, Cluster dendrogram for RNA-seq, showing clustering of samples based on all genes from SSCs, BCSPs and OPs on PODs 5, 10 and 15. **h**, Volcano plot showing differential gene expression between d-SSCs at PODs 10 and 15 and f-SSCs at PODs 10 and 15. Red dots represent genes that are significantly differentially expressed with an adjusted P -value cut-off of 0.05 (DESeq2). Differentially expressed genes that are upregulated in f-SSCs (Fx) are shown in blue (fold change greater than 1.5; P_{adj} less than 0.05), and differentially expressed genes that are upregulated in d-SSCs (DO) are shown in red. **i**, Significantly enriched GO terms for genes upregulated in d-SSCs at PODs 10 and 15, from GREAT version 3.0.0, with P values (one-sided binomial) corrected using the Benjamini–Hochberg correction. **j**, Significantly enriched GO terms for genes upregulated in f-SSCs at PODs 10 and 15 from GREAT, with P values (one-sided binomial) corrected using the Benjamini–Hochberg correction. **k**, As for **h**, but for BCSPs. **l**, As for **i**, but for BCSPs. **m**, As for **j**, but for BCSPs. **n**, As for **h**, but for OPs. **o**, As for **i**, but for OPs. **p**, As for **j**, but for OPs.



Extended Data Fig. 5 | Expression of core signalling genes during distraction. **a**, Heat map showing the expression of Wnt, Notch, bone morphogenetic protein (BMP) and FGF family members, as well as integrins, matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs), in SSCs. Values from RNA-seq in transcripts per million (t.p.m.) are z-scored for each gene to reflect differences across conditions and time. Only genes that are expressed (t.p.m. > 3) in at least one sample are shown. **b**, As for **a**, but for BCSPs. **c**, As for **a**, but for OPs. **d**, Heat map showing expression (in t.p.m.) of Runx-family transcription

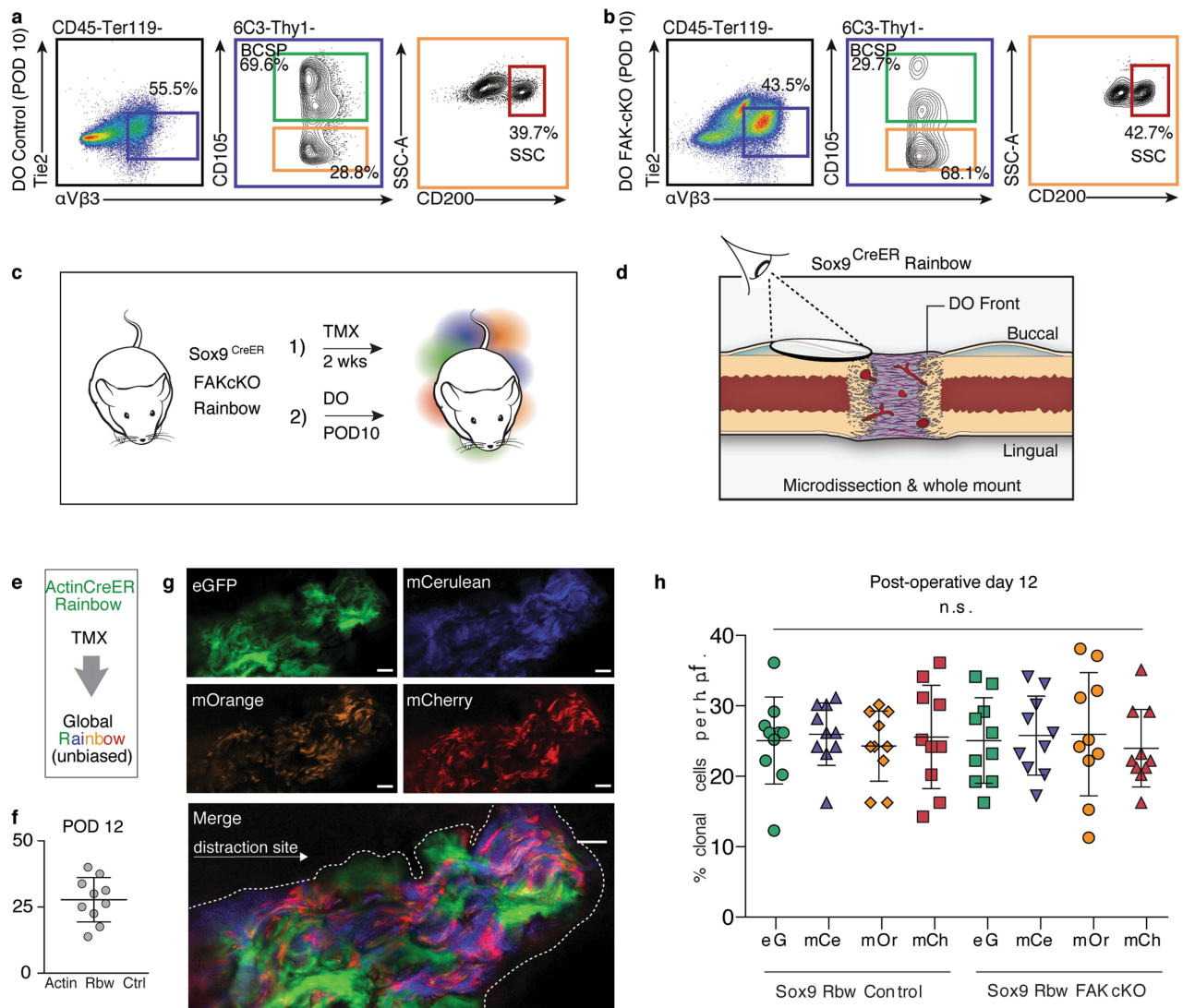
factors in POD5 SSCs, and in POD10 and POD15 f-SSCs and d-SSCs. **e**, As for **d**, but for Dlx-family transcription factors. **f**, Heat map showing accessibility from ATAC-seq in SSCs and BCSPs at Runx2 chromatin immunoprecipitation (ChIP)-seq peaks from MC3T3 preosteoblasts¹²; 6,73 binding sites from ChIP-seq are shown⁴¹. Plotted are the row z-scored normalized counts from SSC and BCSP ATAC-seq data. **g**, As for **f**, but for Dlx5 ChIP-seq data³⁸. ChIP-seq binding sites (24,365) from MC3T3 preosteoblasts are shown.



Extended Data Fig. 6 | See next page for caption.

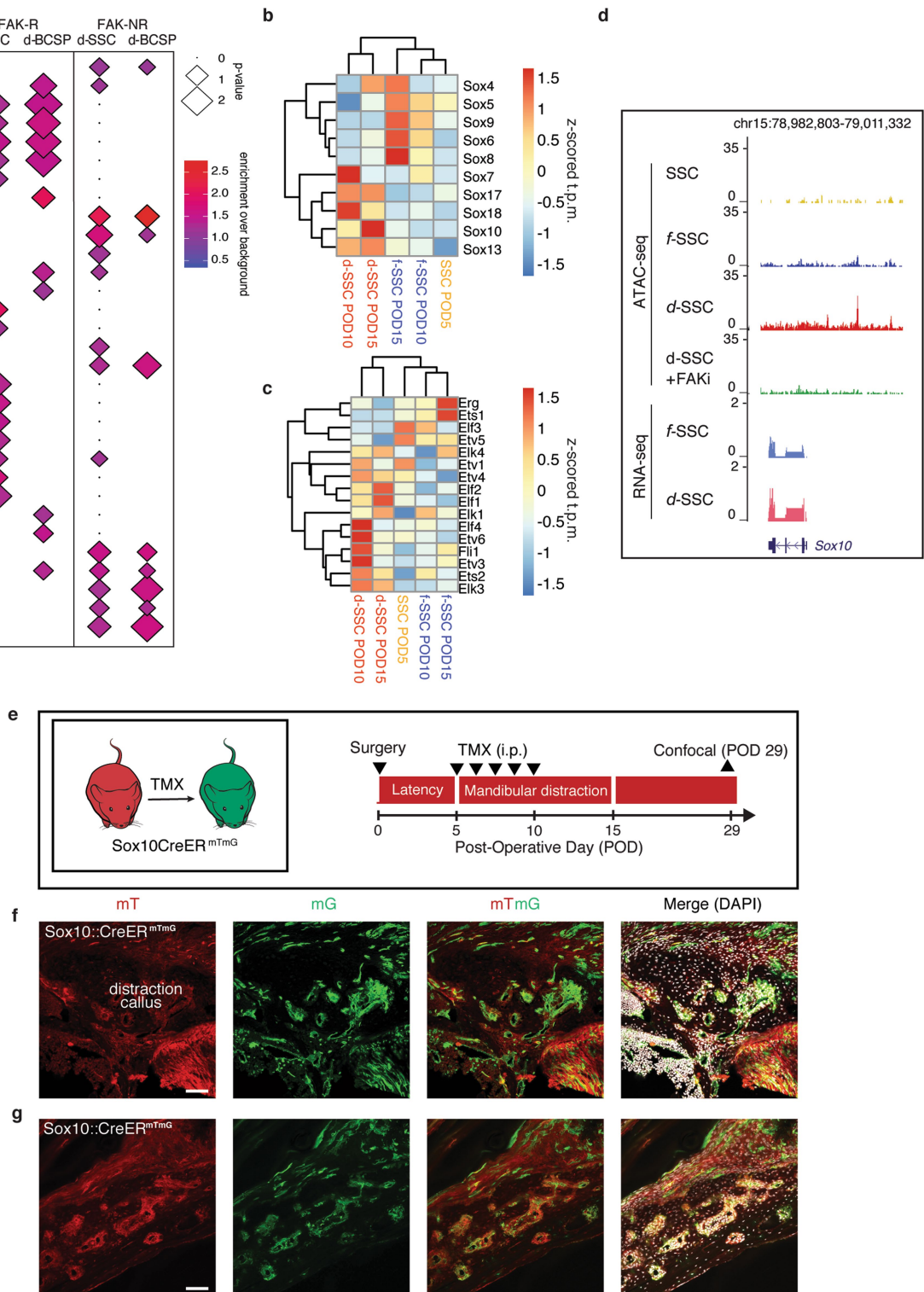
Extended Data Fig. 6 | Expansion of SSCs and progenitor cells depends on FAK. **a**, Timeline for FAK inhibition and conditional genetic knockout during distraction, with respective control conditions, for μ CT analysis (Fig. 3a–d) at mid-consolidation (POD29). **b**, Pentachrome staining of transverse section from mandibles treated with DMSO during distraction osteogenesis and collected at POD15. The white dotted line indicates cortical bone at the osteotomy site; the red dotted line indicates bone that newly formed in response to distraction. Representative of three independent replicates. **c**, As for **b**, but for FAKi treatment. The green dotted line indicates cartilage that newly formed in response to FAKi during distraction osteogenesis. **d**, As for **b**, but for POD29. **e**, As for **c**, but for POD29. **f**, Schematic showing FACS isolation of SSCs and BCSPs from mandible calluses for subsequent in vitro analyses. **g**, **h**, FACS isolation of SSCs and BCSPs from distraction control mandibles receiving DMSO injections (**g**; 55.5%) versus FAK-inhibitor injections (FAKi) (**h**; 41.7%), revealing diminished expansion of the SCC hierarchy (first column, black gate) in response to FAK inhibition by POD10. The ratio of SSCs (red

gate) to their committed bone progenitors (BCSPs, green gate) during distraction was substantially disrupted by FAK inhibition, such that the proportion of SSCs was higher. Representative of three independent replicates. **i**, Quantification of the frequency of d-SSCs, d-BCSPs and d-OPs within mandibular calluses collected at POD10 after FAKi or control injections ($n = 6$; error bars indicate s.d. from the mean). **j**, Representative fluorescence micrographs showing phosphorylated-FAK activity (left column, green) in d-SSCs after treatment with DMSO (top) or with FAK inhibitor (bottom). Right column, phospho-FAK fluorescence merged with DAPI fluorescence. Representative of three independent replicates. **k**, Quantification of d-SSC colony formation in vitro in response to FAKi and in a FAK(cKO), compared with their respective control conditions ($n = 15$ per condition; *** $P < 0.001$, Tukey's multiple comparisons; means \pm s.d.). Direct comparison of colony formation in FAKi and FAK(cKO) conditions was not significantly different. n refers to the number of animals in each independent experiment.



Extended Data Fig. 7 | Cellular dynamics during distraction osteogenesis in FAK(cKO) mice. **a, b**, FACS isolation of cells from control mice (**a**, *Sox9^{creERT2};Ptk2^{fl/fl}*, no TMX) and FAK(cKO) mice (**b**, *Sox9^{creERT2};Ptk2^{fl/fl}*, TMX treatment) shows that the SSC hierarchy (first column, black gate) is disrupted similarly in FAK(cKO) mice and in FAKi (Extended Data Fig. 6h). In the FAK(cKO) mice, the proportion of downstream multipotent BCSPs (green gate) compared with SSCs (orange and red gates) is lower than in controls. Representative of three independent replicates. **c**, Experimental strategy for clonal analysis of FAK(cKO) Rainbow contribution to regeneration in response to distraction, using whole-mount specimens from callus microdissection of mandibles in *Sox9^{creERT2};R26^{Rainbow}* mice. **d**, Vantage point for the acquisition of the confocal images of whole-mount specimens shown in Fig. 3g, h. **e**, Experimental strategy for clonal analysis in

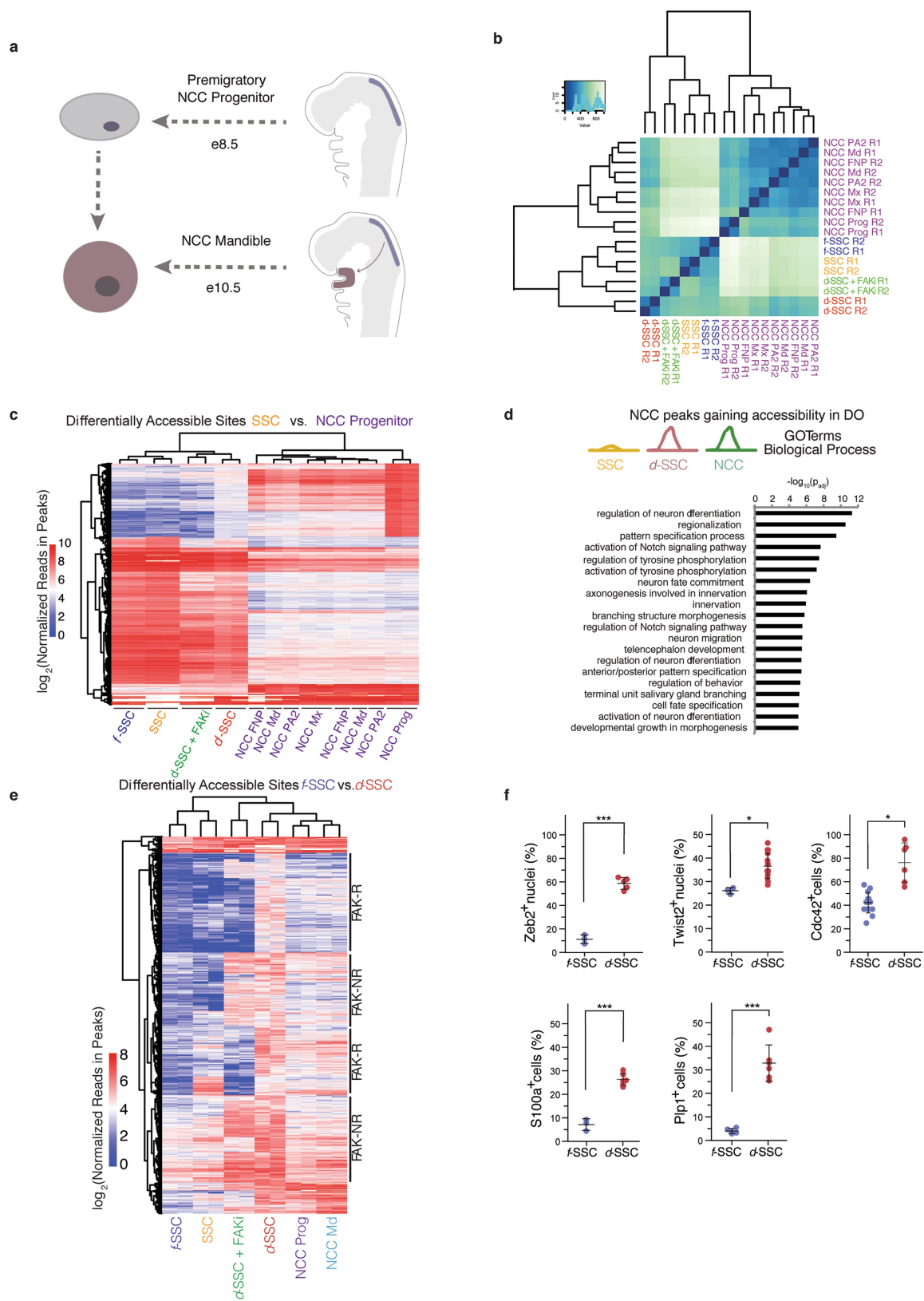
Actin^{creERT2};R26^{Rainbow} mice. **f**, Quantification of average clone size at POD12 during distraction osteogenesis in *Actin^{creERT2};R26^{Rainbow}* mice ($n = 10$; mean \pm s.d.). **g**, Whole-mount imaging at the site of distraction at POD12 in *Actin^{creERT2};R26^{Rainbow}* mice. The view is a lateral-to-medial view of callus overlying the distraction site (indicated by the white dotted outline overlying the distraction gap). The distraction gap contains large clones with a migratory spreading phenotype ($n = 4$ for each of three independent replicates). **h**, Quantification of clones (whole-mount imaging) that are positive for each fluorophore per h.p.f. ($\times 40$ magnification) two weeks after tamoxifen induction (one injection per day, five days total) plus distraction until POD12, using *Sox9^{creERT2};R26^{Rainbow}* mice (left) and FAK(cKO)^{Rainbow} mice (right) ($n = 10$; mean \pm s.d.). n refers to the number of animals in each independent experiment.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Involvement of neural crest transcription factors in SSCs during distraction. **a**, Motifs enriched in FAK-R and FAK-NR sites in SSCs and BCSPs. The size of each diamond represents the negative P value for enrichment (one-sided binomial, Benjamini–Hochberg correction); colours represent the percentage of sites in target/percentage of sites in background. **b**, Heat map showing expression of Sox-family transcription factors. Colours represent z -scored t.p.m. values from RNA-seq at PODs 5, 10 and 15 for d-SSCs and f-SSCs. Only factors with t.p.m. greater than or equal to 3 in at least one sample are shown. **c**, As for **b**, but for Ets-family transcription factors. **d**, The *Sox10* locus (with its location on chromosome 15 shown at the top), showing accessibility at the promoter that is distraction-specific (red), congruent with the expression data in **b** and the RNA-seq track below. The signal for tracks is normalized by read depth and overall peak enrichment in the library. Tracks are

representative of two biological replicates. **e**, Experimental scheme for tamoxifen induction of *Sox10^{creERT2};R26^{mT/mG}* mice during the early and mid-distraction phase (POD5–10) to trace the Sox10⁺ lineage in a distraction-specific context. Mandibles were collected for confocal microscopy at POD29 ($n = 4$). **f**, Confocal micrograph of *Sox10^{creERT2};R26^{mT/mG}* mandible at POD29, demonstrating the capability of Sox10⁺ cells to give rise to the distraction regenerate. Filters are, from left to right, mT (mTomato, background), mG (mGFP, Sox10⁺ lineage), mTmG (merged) and merged (mTmG with DAPI). **g**, Confocal micrograph of *Sox10^{creERT2};R26^{mT/mG}* mandible at POD29, showing the presence of Sox10-lineage cells within surrounding callus and periosteum to give rise to the regenerate. Scale bars, 200 μ m (**b**, **c**). n refers to the number of animals in each independent experiment.

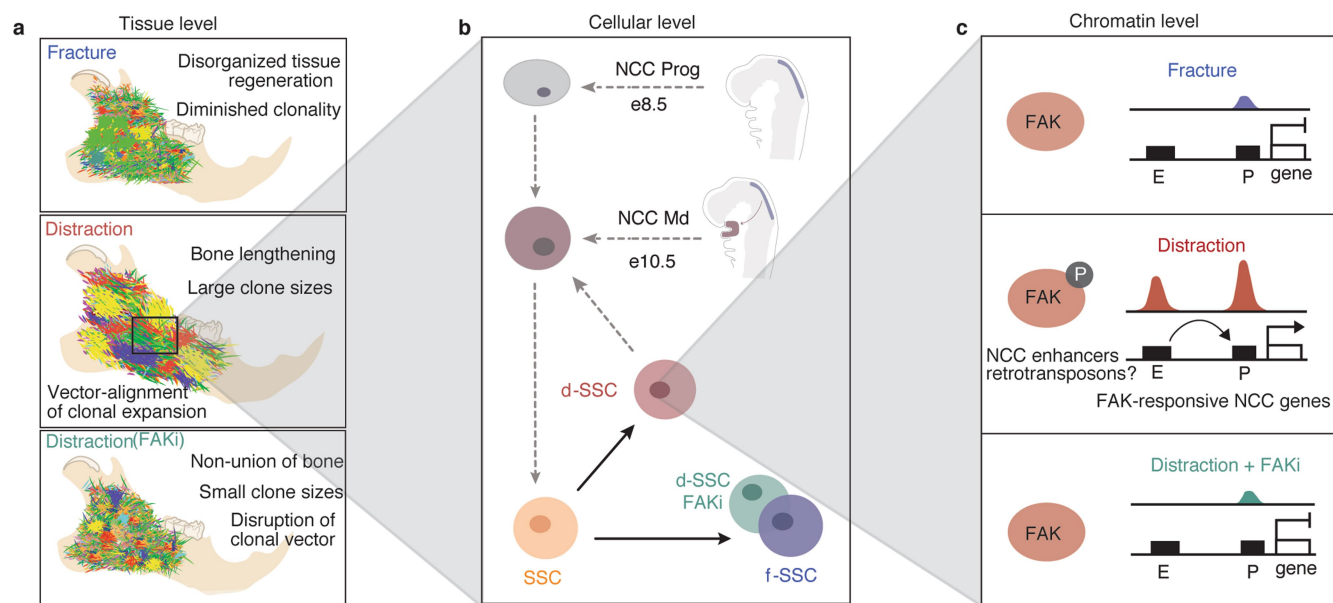


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | NCC transcriptional networks in d-SSCs.

a, Diagram showing the developmental origin of NCC-derived SSCs. **e**, embryonic day of development. **b**, Clustering of SSC and NCC samples using all accessible sites in SSCs and NCCs merged. **c**, Heat map showing all accessible sites that are significantly differentially accessible between uninjured SSCs and NCC progenitors. Colour represents the \log_2 -transformed normalized read counts within peak regions. **d**, GO terms enriched in sets of genes near peaks that are more accessible in NCC and d-SSCs than in uninjured SSCs. GO terms shown are those that are significantly enriched after false discovery rate correction in GREAT

(two-sided binomial P value shown). **e**, Heat map showing all accessible sites that are differentially accessible between f-SSCs and d-SSCs. The sites that are FAK-R or FAK-NR are highlighted along the right. Colour represents the \log_2 -transformed normalized read counts within peak regions. **f**, Quantification of immunofluorescence staining for each of the NCC markers observed on RNA-seq analysis and evaluated in this figure, including nuclear Zeb2 ($n = 15$, $***P < 0.001$), nuclear Twist2 ($n = 15$, $*P < 0.05$), Cdc42 ($n = 6$, $P < 0.05$), S100a ($n = 6$, $***P < 0.001$) and Plp1 ($n = 6$, $***P < 0.001$) (Student's t -test; shown are means \pm s.d.). n refers to the number of animals in each independent experiment.



Extended Data Fig. 10 | Controlled mechanical advancement of the lower jaw unlocks neural crest potential for regeneration of the mandible. **a**, At the tissue level, clonality within the mandible during distraction (middle) is observed to occur in a highly linear and directional manner in parallel to the vector of distraction. By contrast, clonality observed in fracture (top) and distraction plus FAKi (bottom) was highly mesenchymal with less apparent organization in its morphology, indicating a nondirectional clonal proliferation. **b**, The cellular level: the developmental origin of the NCC-derived SSCs of the mandible (top), and the postnatal SSCs of the mandible (bottom) that are present in our experiments. During distraction the d-SSC (shown in pink) demonstrates plasticity and takes on an NCC-derived signature, whereas the f-SSC (purple) retains its postnatal SSC characteristics with no NCC signature. In the absence of FAK signalling, the d-SSC FAKi (green) reverts functionally and epigenomically to the fracture state without emergence

of the NCC signature. 'NCC Prog' indicates the premigratory (e8.5) NCC progenitor population; 'NCC Md' indicates the postmigratory (e10.5) NCC population arriving within the mandible. **c**, At the chromatin level, distraction induces a gain in accessibility at promoters (P) of FAK-responsive NCC craniofacial genes through the activation of their enhancers (E), with a parallel gain in accessibility of retroviral elements near NCC-specific craniofacial enhancers. Thus mechanotransduction during mandible distraction unlocks FAK-responsive craniofacial enhancers, potentially through retrotransposons, enacting a developmental NCC program in d-SSCs similar to that of the e10.5 NCC Md population (**b**). This does not occur under fracture conditions (top) or during distraction plus FAKi (bottom). These differential epigenomic responses correlate with the degree of clonality and patterning seen in Rainbow mice (**a**) that occurs in response to distraction. Circled P represents phosphorylation.

DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells

Yizhou Joseph He¹, Khyati Meghani¹, Marie-Christine Caron^{2,3}, Chunyu Yang¹, Daryl A. Ronato^{2,3}, Jie Bian¹, Anchal Sharma⁴, Jessica Moore¹, Joshi Niraj¹, Alexandre Detappe⁵, John G. Doench⁶, Gaelle Legube⁷, David E. Root⁶, Alan D. D'Andrea^{1,8}, Pascal Drané¹, Subhajyoti De⁴, Panagiotis A. Konstantinopoulos⁵, Jean-Yves Masson^{2,3} & Dipanjan Chowdhury^{1,6,9*}

Limited DNA end resection is the key to impaired homologous recombination in *BRCA1*-mutant cancer cells. Here, using a loss-of-function CRISPR screen, we identify DYNLL1 as an inhibitor of DNA end resection. The loss of DYNLL1 enables DNA end resection and restores homologous recombination in *BRCA1*-mutant cells, thereby inducing resistance to platinum drugs and inhibitors of poly(ADP-ribose) polymerase. Low *BRCA1* expression correlates with increased chromosomal aberrations in primary ovarian carcinomas, and the junction sequences of somatic structural variants indicate diminished homologous recombination. Concurrent decreases in DYNLL1 expression in carcinomas with low *BRCA1* expression reduced genomic alterations and increased homology at lesions. In cells, DYNLL1 limits nucleolytic degradation of DNA ends by associating with the DNA end-resection machinery (MRN complex, BLM helicase and DNA2 endonuclease). In vitro, DYNLL1 binds directly to MRE11 to limit its end-resection activity. Therefore, we infer that DYNLL1 is an important anti-resection factor that influences genomic stability and responses to DNA-damaging chemotherapy.

Patients with high-grade serous ovarian carcinoma (HGSOC) and germline mutations in *BRCA1* and *BRCA2* exhibit high sensitivity and improved outcome to double-strand DNA break (DSB)-inducing agents (that is, platinum and inhibitors of poly(ADP-ribose) polymerase (PARP)) owing to underlying defects in DNA repair via homologous recombination^{1–3}. Owing to their effectiveness, three PARP inhibitors (PARPi; olaparib, rucaparib and niraparib) have recently gained FDA approval for the treatment of HGSOCs. However, de novo and acquired resistance to these agents is common even in *BRCA* mutation carriers, and pose a considerable, and unsolved, clinical challenge. Restoration of homologous recombination by reinstating DNA end resection (for example, by depletion or deletion of 53BP1^{4,5} and interactors^{6,7}, or of REV7⁸ and interactors^{9–13}) in *BRCA1*-mutant cells is sufficient to confer resistance to PARPi. Stabilizing the DNA replication fork (for example, by depletion or deletion of factors such as PTIP¹⁴, CHD4¹⁵ and EZH2¹⁶ that recruit the nucleases MRE11^{17,18}, EXO1¹⁸ and MUS81) in *BRCA1/2*-mutant cells also causes resistance to PARPi. Here we adopted a systematic approach to identify unexplored factors or pathways that could be responsible for resistance to PARPi or platinum therapy in *BRCA*-defective patients with HGSOC.

We used a genome-scale bacterial clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 knockout (GeCKO) library¹⁹ to identify genes in which loss confers resistance to clinical PARPi and platinum drugs in a panel of patient-derived *BRCA1*-mutant HGSOC lines. Among the most notable ‘hits’ of our screen, in both PARPi- and platinum-treated cells, was dynein light chain 1 protein (DYNLL1; also known as LC8 or PIN). Diminished expression of DYNLL1 significantly ($P < 0.01$) correlated with poor progression-free survival (PFS), after platinum-based chemotherapy of patients with *BRCA1*-mutated ovarian carcinomas. We describe the mechanism by which DYNLL1 contributes to resistance to PARPi or platinum therapy.

We report that it inhibits resection by directly interacting with MRE11 that is involved in nucleolytic degradation of DSB ends. These observations depict a previously uncharacterized function of DYNLL1 in regulation of homologous recombination-mediated DSB repair, which bears considerable clinical relevance.

ATMIN and DYNLL1 affect PARPi and platinum sensitivity

We used the updated GeCKO library to screen a panel of *BRCA1*-mutant HGSOC lines (UWB1.289, COV362 and JHOS-2) to identify PARPi- or cisplatin-resistant clones. Specifically, Cas9-expressing cells were infected with single-guide RNA (sgRNA) library, passaged to allow genomic editing, and treated with olaparib or cisplatin for 14 days (Fig. 1a). Olaparib- or cisplatin-resistant clones were analysed using a barcode-based sequencing strategy^{19,20}, which includes the STARS algorithm to identify the disrupted gene. Deep-sequencing results from the resistant clones were normalized against input abundance (representative examples, Extended Data Fig. 1a) and a rank order based on the STARS score and significant enrichment of sgRNAs was generated ($P < 0.01$, Supplementary Tables 1 and 2). As anticipated in the olaparib-resistance screen, factors such as PARP1, 53BP1^{4,5} and the recently described shieldin complex proteins^{9–13} were among the high-confidence hits. In the cisplatin-resistance screen, the membrane channel protein LRRC8D²¹ that has been implicated in the take up of cisplatin was a high-confidence ‘hit’. DYNLL1 was among the top-ranked factors in both the olaparib and platinum screen, as was the transcriptional regulator of DYNLL1, the ATM-interacting protein ATMIN.

DYNLL1 has been implicated as a master regulator of several cellular functions ranging from intracellular trafficking to apoptosis^{22,23}, but its function in DNA repair remains unexplored. We selected ATMIN and DYNLL1 for further study because of their potential clinical relevance.

¹Division of Radiation and Genome Stability, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. ²Genome Stability Laboratory, CHU de Québec Research Center, HDQ Pavilion, Oncology Axis, Québec City, Québec, Canada. ³Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University Cancer Research Center, Québec City, Québec, Canada. ⁴Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA. ⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. ⁶Broad Institute of Harvard and MIT, Cambridge, MA, USA. ⁷LBCMCP, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse, UT3, Toulouse, France. ⁸Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. ⁹Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. *e-mail: dipanjan_chowdhury@dfci.harvard.edu

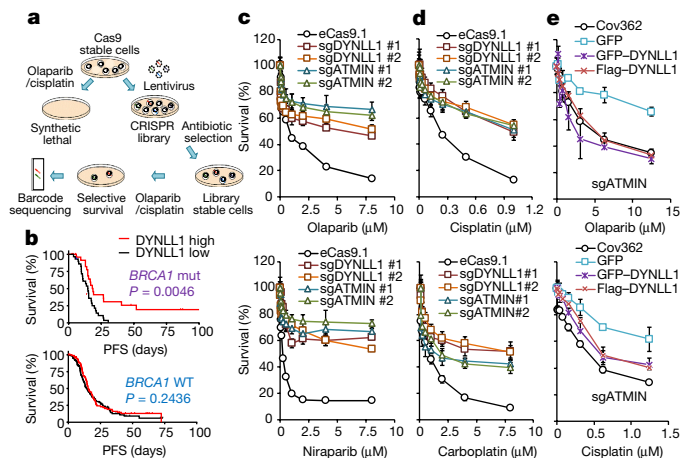


Fig. 1 | Genome-wide CRISPR screen reveals DYNLL1 loss causes resistance to PARPi and platinum in BRCA1-mutant HGSOCs.

a, Schematic of CRISPR-based screen for resistance to PARPi (olaparib) or platinum-based chemotherapy (cisplatin). **b**, PFS of patients with BRCA1-mutant (top, DYNLL1-high expression $n = 16$, DYNLL1-low expression $n = 20$) and BRCA1 wild-type (WT) (bottom, DYNLL1-high $n = 135$, DYNLL1-low $n = 111$) ovarian carcinoma based on above or below median expression values of DYNLL1 (source: ovarian cancer, TCGA dataset²⁴). Statistical significance was assessed by one-sided Mantel–Cox test. **c**, **d**, Survival assays of Cov362 cells after treatment with PARP inhibitors (olaparib or niraparib) (**c**) or platinum drugs (cisplatin or carboplatin) (**d**) following the loss of DYNLL1 or ATMIN via CRISPR knockout sgRNA clones (sgDYNLL1 and sgATMIN). #1 and #2 are independent stable clones; eCas9.1 denotes the vector control. Experiments were repeated three times independently with similar results. **e**, Survival assay of Cov362 ATMIN^{-/-} clone (sgATMIN) expressing tagged DYNLL1, treated with olaparib (top) or cisplatin (bottom). Data are mean \pm s.e.m. from three different experiments.

ATMIN is altered in approximately 30% of HGSOCs (Extended Data Fig. 1b). Low expression of DYNLL1 significantly correlates with worse PFS of BRCA1-mutant patients with HGSOC from The Cancer Genome Atlas (TCGA) dataset²⁴ (Fig. 1b). However, no such correlation was observed with the PFS of BRCA-proficient patients with HGSOC. Furthermore, there is no effect of DYNLL1 depletion on PARPi and platinum sensitivity in BRCA-proficient cells (Extended Data Fig. 1c). This is consistent with the notion that the decreased expression of DYNLL1 may cause platinum resistance exclusively in

BRCA1-mutant patients with HGSOC. In ATMIN- and DYNLL1-deficient Cov362 clones (Extended Data Fig. 1d), we observed that the loss of ATMIN leads to depletion of DYNLL1. However, the reverse was not observed. This is consistent with the role of ATMIN in regulating DYNLL1 expression²⁵. Next, we confirmed that loss of ATMIN or DYNLL1 leads to significant resistance to PARPi drugs, olaparib and niraparib (Fig. 1c), and the platinum drugs, cisplatin and carboplatin (Fig. 1d), in independent clones. We validated these results in another BRCA1-mutant line ovarian line, UWB1.289, BRCA1-mutant line breast lines MDA-MB-436 and L56Br-C1, and by co-depletion of BRCA1 and DYNLL1 in RPE1 and HeLa cells (Extended Data Figs. 1e–h, 2a–c). We conclude that the effect of DYNLL1 depletion on PARPi and platinum drugs in BRCA1-deficient contexts is not lineage-specific. Next, we expressed green fluorescent protein (GFP)- or Flag-tagged DYNLL1 in ATMIN^{-/-} and DYNLL1^{-/-} cells. Restoration of DYNLL1 expression (Extended Data Fig. 2d) in DYNLL1^{-/-} cells restored sensitivity to cisplatin and olaparib (Extended Data Fig. 2e). Importantly, restoration of DYNLL1 expression in ATMIN^{-/-} cells (Extended Data Fig. 2d) also restored sensitivity to cisplatin and olaparib (Fig. 1e). These results confirm the epistatic relationship of ATMIN and DYNLL1, and suggest that the effect of ATMIN loss in resistance to PARPi and platinum in BRCA1-mutant cells is mediated by DYNLL1.

DYNLL1 inhibits DNA end resection

DYNLL1 depletion did not affect the PARPi and platinum sensitivity of BRCA2-mutant cells (KURAMOCHI) (Extended Data Fig. 2f), and low levels of DYNLL1 do not correlate with worse PFS in BRCA2-mutant patients with HGSOC (Extended Data Fig. 2g). Therefore, we hypothesized that DYNLL1 may function upstream of the BRCA2-mediated RAD51 nucleofilament and restore homologous recombination to induce PARPi resistance in BRCA1-mutant cells. Consistent with our hypothesis, we observed a notable increase in the formation of RAD51 foci in DYNLL1^{-/-} BRCA1-mutated Cov362 cells treated with olaparib (Fig. 2a) or exposed to ionizing radiation (Extended Data Fig. 3a). To rule out the possibility that mutant BRCA1 is involved in PARPi resistance in DYNLL1-deficient cells, we silenced BRCA1 expression in Cov362 cells, and confirmed that DYNLL1 loss restores RAD51 foci in the absence of BRCA1 (Extended Data Fig. 3b).

Loss of 53BP1 foci formation in BRCA1-deficient tumours restores homologous recombination and causes resistance to PARPi. DYNLL1 interacts with 53BP1²⁶; therefore, we examined the efficacy of 53BP1 foci formation in DYNLL1^{-/-} cells. The number of 53BP1 foci per

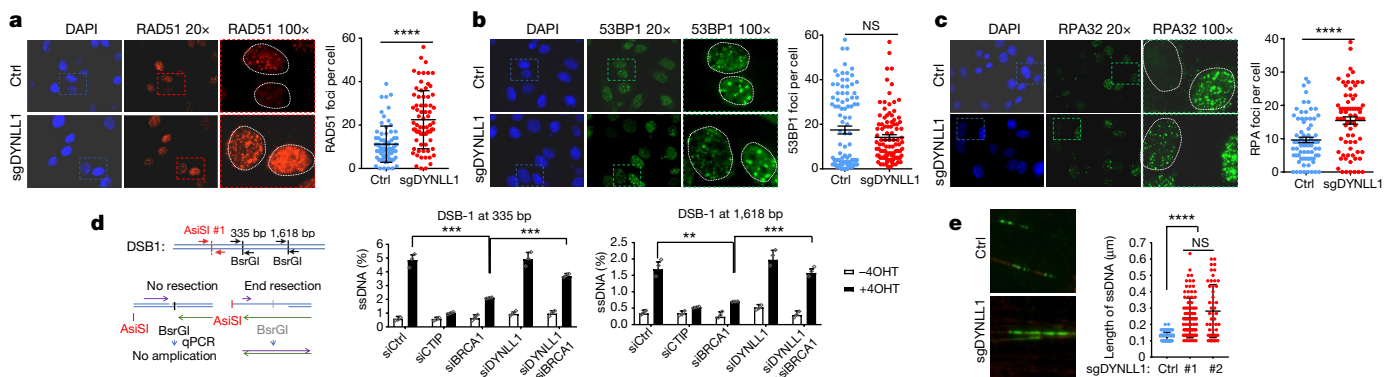


Fig. 2 | DYNLL1 loss leads to restoration of DNA end resection and homologous recombination. **a–c**, Immunofluorescence and quantification of control (ctrl) and DYNLL1^{-/-} (sgDYNLL1) Cov362 cells treated with 10 μ M olaparib for 24 h and stained with antibodies against RAD51 (**a**; 143 cells), 53BP1 (**b**; 102 cells) and RPA32 (**c**; 146 cells). NS, not significant. **** $P < 0.0001$, unpaired two-tailed Student's t -test. γ -H2AX staining was performed for all panels (data not shown). **d**, Left, schematic of AsiSI-based DNA end-resection assay in U2OS cells. Right, PCR-based quantification of ssDNA formation at 335 or 1,618 bp downstream of the

AsiSI-induced break site in cells transfected with the indicated siRNAs, and with or without 4-hydroxytamoxifen (4OHT) ($n = 3$). ** $P < 0.01$, *** $P < 0.001$, unpaired two-tailed Student's t -test. **e**, Change in ssDNA length in control and DYNLL1^{-/-} (sgDYNLL1) Cov362 cells after treatment with 10 μ M olaparib for 48 h. Representative images (left) and quantification (right). In total, 150 DNA fibres were analysed. **** $P < 0.0001$, unpaired two-tailed Student's t -test. NS, $P = 0.0848$. Data are mean \pm s.e.m. from three independent experiments.

cell is not significantly changed in the absence of DYNLL1 (Fig. 2b). However, there is a decrease in cells with relatively high numbers (more than 30) of 53BP1 foci (Extended Data Fig. 3e). The modest effect of DYNLL1 loss on 53BP1 foci formation is unlikely to account for the PARPi-resistance phenotype. Next, we examined DNA end resection in DYNLL1-deficient cells. Relative to control Cov362 cells, there was a significant increase in the formation of RPA foci in *DYNLL1*^{-/-} Cov362 cells after treatment with olaparib (Fig. 2c) or exposure to ionizing radiation (Extended Data Fig. 3c). Also, levels of phosphorylated RPA32 were increased after olaparib treatment in *DYNLL1*^{-/-} Cov362 cells relative to control Cov362 cells, suggesting an increase in end resection (Extended Data Fig. 3d). In BRCA1-proficient RPE1 cells, DYNLL1 depletion had no effect on the number of RAD51 and 53BP1 foci per cell, and there was a modest increase in the number of RPA32 foci per cell (Extended Data Fig. 3f–h).

To measure DNA end resection quantitatively, we introduced the AsiSI endonuclease fused to the oestrogen receptor (ER-AsiSI)²⁷ in either BRCA1-depleted or DYNLL1- and BRCA1-depleted U2OS cells, and used a quantitative PCR (qPCR)-based method²⁸ to measure single-stranded DNA (ssDNA) (Fig. 2d). We observed a decrease in the amount of ssDNA generated from a specific DSB after depletion of BRCA1. However, when we silenced both DYNLL1 and BRCA1, we ‘rescued’ this phenotype and ssDNA levels were restored (Fig. 2d). Finally, we adopted a high-resolution method, single molecule analysis of resection tracks (SMART)²⁹, to visualize and measure ssDNA generated by DNA end resection in the *BRCA1*-mutant Cov362 cells after olaparib treatment. The loss of DYNLL1 enhanced end resection in these cells and significantly increased the production of ssDNA after olaparib treatment (Fig. 2e). Together, our results suggest that the loss of DYNLL1 in *BRCA1*-mutant or -depleted cells restores the formation of RAD51 foci by facilitating DNA end resection.

DYNLL1 influences genome stability of primary HGSOCS

To investigate any connection of BRCA1 and DYNLL1 in the genomic stability of primary HGSOCS, we analysed data from a cohort of 112 tumours from 92 patients with HGSOCS³⁰. Whole-genome and transcriptome sequencing had revealed approximately 36,000 somatic structural variants (amplifications, deletions, translocations and inversions) in these tumours. Because the loss of DYNLL1 leads to restoration of homologous recombination-mediated DSB repair in *BRCA1*-mutant cells, we compared patterns of structural variants in this cohort, after grouping the samples into four categories (Fig. 3a, b) based on combinatorial high (above median) and low (below median) expression of BRCA1 and DYNLL1. Compared to the BRCA1-high group, the BRCA1-low group had significantly ($P < 0.001$) higher number of structural variants per genome, which is consistent with the concept of increased genomic instability in BRCA1-deficient tumours. However, in the samples that had low expression of both BRCA1 and DYNLL1, the number of structural variants was significantly lower than the BRCA1-low group ($P = 0.01$), which supports our experimental observation that DYNLL1 loss facilitates restoration of DNA repair in a BRCA1-deficient background. By analysing different classes of structural variants (that is, deletions, duplications, insertions and translocations), we observed broadly similar patterns (increased burden of genomic aberrations in the BRCA1-low group, which is reduced in the group with low expression of both BRCA1 and DYNLL1). Nonetheless, these samples still had a relatively increased burden of structural variants, particularly duplications and deletions ($P < 0.05$), compared to the DYNLL1- and BRCA1-proficient samples (Fig. 3a). Next, we analysed the extent of base-pair level homology at the structural variant junctions for these samples. In samples with low and high expression of BRCA1 and DYNLL1, respectively, the proportion of structural variants with indications of homologous recombination-mediated repair (± 10 -bp homology) was smaller than the groups with high expression of BRCA1 ($P < 0.05$ in both cases). By contrast, in the group with low expression of both BRCA1 and DYNLL1, the proportion of structural variants with indications of homologous recombination-mediated

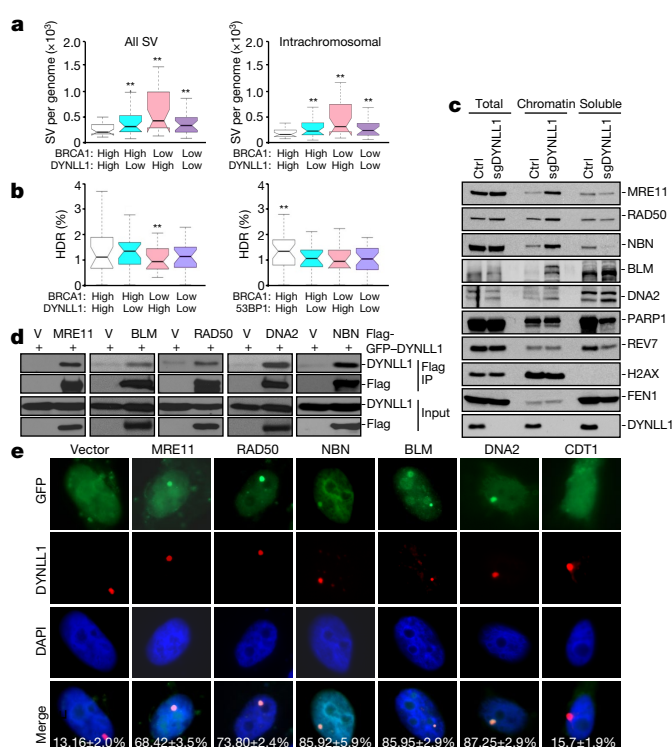


Fig. 3 | Effect of DYNLL1 on chromosomal aberrations in HGSOCS samples and interaction with the DNA end-resection machinery. **a**, Samples from the Australian Ovarian Cancer Study (OV-AU) cohort were grouped based on combinatorial expression levels of BRCA1 and DYNLL1 (median cut-off). The frequencies of somatic structural variations (SV) (left) and of intrachromosomal structural variations (right) were plotted. **b**, Samples from the OV-AU cohort were grouped based on the combinatorial expression levels of BRCA1 and DYNLL1 (left, median cut-off) or BRCA1 and 53BP1 (right, median cut-off). The frequency of structural variants (deletions, duplication, insertions, intrachromosomal translocation) with indications of homologous recombination (HDR; ≥ 10 -bp homology) was plotted. **c**, Representative immunoblots from three experiments of indicated proteins in control and *DYNLL1*^{-/-} (sgDYNLL1) Cov362 cells after subcellular fractionation. Quantification ($n = 3$) of chromatin enrichment is shown in Extended Data Fig. 3e. **d**, Flag immunoprecipitation (IP) of GFP-DYNLL1 with Flag-tagged DNA end-resection proteins. Experiments were repeated four times independently with similar results. V, vector control. **e**, U2OS stable clone (U2OS19) containing 256 lac-operator and 96 tetracycline response element copies were transiently transfected with mCherry-LacI fused to DYNLL1 and GFP either alone or fused to the indicated proteins. Quantification of the co-localization of mCherry-DYNLL1 with the indicated GFP-tagged proteins on the LacO array is indicated (values denote mean \pm s.d., $n = 3$). Original magnification, 100 \times .

repair (≥ 10 -bp homology) had increased and was comparable to that observed in the homologous recombination-proficient group with high expression of both BRCA1 and DYNLL1. Analysis with high and low 53BP1 yielded a very similar pattern with low 53BP1 restoring homology in BRCA1-low samples (Fig. 3b). This is consistent with the role of 53BP1 in impeding resection in *BRCA1*-mutant cells⁵. Together, the analysis of genomic alterations in HGSOCS provides independent evidence that the suppression of DYNLL1 expression distinctly compensates for BRCA1 deficiency in restoring DNA repair functions and maintaining genomic stability. Furthermore, correlation of diminished DYNLL1 expression with increased homology at structural variants in the BRCA1-low cohort suggests that this decrease in DYNLL1 probably occurs concurrently with BRCA1 loss otherwise the genomic scar at a structural variant would represent other DNA repair mechanisms. We also obtained structural variant and expression data for cohorts from the Pancreatic Cancer Endocrine neoplasms (PAEN-AU), which are part of the International Cancer Genome Consortium (ICGC).

There were 4–154 (median: 18) structural variants per sample in the PAEN-AU cohort. Extending the analyses to this pancreatic cancer cohort, we observed that in the samples with low expression of both BRCA1 and DYNLL1, the number of structural variants was lower ($P = 0.02$, Extended Data Fig. 4a), and the proportion of structural variants with indications of homologous recombination-mediated repair (≥ 10 bp homology) was slightly higher, but not statistically significant ($P > 0.05$), than those with low expression of only BRCA1 (Extended Data Fig. 4b), which is consistent with the patterns in the ovarian cancer cohort and our experimental observations.

DYNLL1 associates with DNA end-resection machinery

Next, we investigated the molecular mechanism of how DYNLL1 affects DNA end resection. ATMIN and DYNLL1 individually and as a complex have been implicated in the regulation of transcription^{25,26,31}. Analysis of the RNA-sequencing data from *Atmin*^{-/-} mouse embryonic fibroblasts revealed an increase in the mRNA levels of the nucleases *Dna2* and *Mre11a* and the Bloom's syndrome helicase (*Blm*) after aphidicolin treatment³². Following up on this data, we observed a basal increase in the transcripts of DNA end-resection factors in ATMIN- and DYNLL1-deficient Cov362 cells (Extended Data Fig. 4c, d), but this moderate increase in transcripts did not translate to a consistent increase in the protein levels of all the factors (Fig. 3c). However, subcellular fractionation of these cells revealed that loss of DYNLL1 leads to very distinct increase in the DNA end-resection factors in the chromatin-bound fraction (Fig. 3c, Extended Data Fig. 4e). Unrelated DNA repair factors such as REV7, PARP1 and FEN1 are not enriched in the chromatin of DYNLL1-deficient cells. DYNLL1 is an essential 'hub' protein that interacts with hundreds of proteins^{33,34} and potentially influences their function. For example, it inhibits the activity of the apoptotic factors BIM and BMF³⁵ and nitric oxide synthase (NOS) via interaction³⁶. Therefore, we tested the association of DYNLL1 with the DNA end-resection factors and observed that DYNLL1 interacts with Flag-tagged (Fig. 3d) and endogenous MRE11, RAD50, NBN, DNA2 and BLM (Extended Data Fig. 4f). To investigate these interactions in intact cells further, we adopted the *Escherichia coli* LacI/LacO tethering system³⁷. mCherry-DYNLL1 was fused to the lac-repressor (LacI) and its co-localization with the GFP-tagged DNA end-resection factors was observed in approximately 70% of cells that were analysed (Fig. 3e). Therefore, we concluded that DYNLL1 associates with the DNA end-resection machinery (MRE11, NBN, RAD50, BLM and DNA2). The interaction of DYNLL1 with these DSB factors is consistent with the observation that DYNLL1 is recruited to DSBs (Extended Data Fig. 4g) and overall there is more chromatin-associated DYNLL1 after DNA damage (Extended Data Fig. 4h).

DYNLL1 and end-resection factors mediate PARPi response

DYNLL1 is present in both monomeric and dimeric forms in cells, and the structural transition³⁸ (Protein Data Bank (PDB) accession 3DVT, Extended Data Fig. 5a) between these two forms might be key to its interaction network³⁴. Phosphorylation of Ser88 is crucial for dimerization and ability to interact with factors such as BIM³⁹. Another residue, Cys2, in the disordered region of the protein has also been implicated in interactions⁴⁰. The phospho-null mutation of Ser88 (Ser88Ala) or mutation of Cys2 to alanine (Cys2Ala) disrupted the interaction of DYNLL1 with the DNA end-resection proteins. By contrast, the phosphomimetic Ser88 mutation (Ser88Asp) had a detectably enhanced interaction with MRE11, NBN, RAD50, BLM and DNA2 (Fig. 4a). Notably, the well-characterized interaction of DYNLL1 with 53BP1³⁴ had a contrasting pattern, with a reduced interaction of 53BP1 with the Ser88Asp mutant and a continued interaction with the Ser88Ala mutant (Extended Data Fig. 5b). Reconstituting *BRCA1*-mutant DYNLL1-deficient cells with wild-type or mutant DYNLL allowed us to determine whether the interaction of DYNLL1 with these factors is regulating the sensitivity of *BRCA1*-mutant cells to PARPi. Wild-type or Ser88Asp mutant DYNLL1 rescued the PARPi-resistance phenotype, whereas the two mutants (Ser88Ala and Cys2Ala) that do not interact

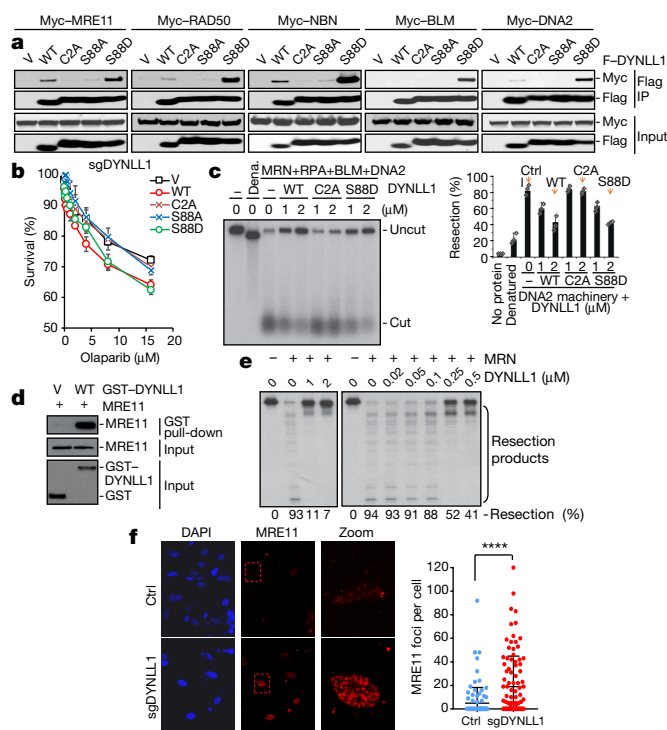


Fig. 4 | Identification and characterization of DYNLL1 mutants that affect genome stability in cells and DNA end resection in vitro.

a, Immunoprecipitation of DYNLL1 mutants with end-resection complex enzymes. Experiments were repeated three times independently with similar results. **b**, Survival assay after olaparib treatment. Data are mean \pm s.e.m. from three different experiments. **c**, Resection products of wild-type (WT) or mutant recombinant DYNLL1 with MRN-RPA-BLM-DNA2 and a ³²P-labelled linear 2.7-kb double-stranded DNA (dsDNA) substrate (left) and quantification of resection efficiency (right) ($n = 3$). Dena., denatured. **d**, Glutathione S-transferase (GST) pull-down of GST-tagged DYNLL1 and MRE11 isolated from insect cells. **e**, Resection product of a 5'-end labelled 100-bp dsDNA incubated with indicated concentration of purified recombinant human DYNLL1 and the MRE11-RAD50-NBS1 (MRN) complex. **f**, Immunofluorescence and quantification of MRE11 foci in wild-type and *DYNLL1*^{-/-} Cov362 cells treated with 10 μ M olaparib for 24 h. Original magnification, 20 \times for DAPI and MRE11, 100 \times for zoom. A total of 198 cells were analysed. **** $P < 0.0001$, unpaired two-tailed Student's *t*-test. Data are mean \pm s.e.m. from three different experiments.

with DNA end-resection enzymes phenocopied DYNLL1-deficient cells (Fig. 4b, Extended Data Fig. 5c).

DYNLL1-MRE11 interaction inhibits DNA end resection

Next, we tested whether DYNLL1 can inhibit DNA end resection in a cell-free system. Purified MRN complex, RPA, DNA2 and BLM was used to conduct in vitro end-resection assays with radiolabelled DNA as previously described⁴¹. Consistent with their ability to interact with the DNA end-resection machinery in *BRCA1*-mutant cells and induce resistance to PARPi, recombinant wild-type DYNLL1 and the Ser88Asp mutant suppressed the end-resection activity of the MRN complex in the presence of BLM, DNA2 and RPA (Fig. 4c). By contrast, the addition of Cys2Ala (Fig. 4c) or Ser88Ala (Extended Data Fig. 5d) mutant proteins that fail to interact with DNA end-resection factors had a limited effect on end-resection activity. Using in vitro binding assays, we only detected the interaction of MRE11 with wild-type DYNLL1 and Ser88Asp mutant proteins (Fig. 4d); that is, other factors included in the in vitro resection assays (DNA2, BLM, EXO1 and RPA32) did not interact (Extended Data Fig. 5e). Therefore, we conducted in vitro end-resection assays excluding these factors and only using the MRN complex. DYNLL1 suppressed the end-resection activity of the MRN complex (Fig. 4e) and did not affect the helicase activity of BLM (Extended Data Fig. 5f). Finally, focusing on the effect of DYNLL1

on MRE11 in cells, we observed that relative to control Cov362 cells, there was a significant increase in the formation of MRE11 foci in *DYNLL1*^{-/-} Cov362 cells after treatment with olaparib (Fig. 4f).

Discussion

The potent end-resection activity of the MRN complex needs to be stringently regulated for efficient DSB repair and to maintain genomic stability. Our results definitively describe a role for *DYNLL1* in the negative regulation of DNA end resection. The primary mechanism by which *DYNLL1* impairs end resection in *BRCA1*-deficient cells is via its interaction with MRE11. How the interaction of *DYNLL1* with MRE11 impairs its nuclease activity or its recruitment to foci remains unexplored. Furthermore, how this interaction is regulated in the context of the DNA damage response, for example, or which kinase or phosphatase(s) regulates the phosphorylation of Ser88 of *DYNLL1* to modulate the interaction needs to be investigated. Another key aspect is that the role of *DYNLL1* in regulating DSB repair is only manifested in *BRCA1*-deficient and not *BRCA2*-deficient cells. This highlights the role of *BRCA1* in the initial end-resection step of homologous recombination, which is independent of *BRCA2*. Importantly, this is very similar to the effect of 53BP1, which was also observed only in *BRCA1*-deficient cells^{4,5}. Future studies will reveal the regulation of this intriguing small protein, which functions inside preformed complexes and connects diverse biological modules.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0670-5>.

Received: 28 April 2018; Accepted: 25 September 2018;

Published online 31 October 2018.

- Bolton, K. L. et al. Association between *BRCA1* and *BRCA2* mutations and survival in women with invasive epithelial ovarian cancer. *JAMA* **307**, 382–390 (2012).
- Vencken, P. M. et al. Chemosensitivity and outcome of *BRCA1*- and *BRCA2*-associated ovarian cancer patients after first-line chemotherapy compared with sporadic ovarian cancer patients. *Ann. Oncol.* **22**, 1346–1352 (2011).
- Yang, D. et al. Association of *BRCA1* and *BRCA2* mutations with survival, chemotherapy sensitivity, and gene mutator phenotype in patients with ovarian cancer. *JAMA* **306**, 1557–1565 (2011).
- Bouwman, P. et al. 53BP1 loss rescues *BRCA1* deficiency and is associated with triple-negative and *BRCA*-mutated breast cancers. *Nat. Struct. Mol. Biol.* **17**, 688–695 (2010).
- Bunting, S. F. et al. 53BP1 inhibits homologous recombination in *Brca1*-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243–254 (2010).
- Callen, E. et al. 53BP1 mediates productive and mutagenic DNA repair through distinct phosphoprotein interactions. *Cell* **153**, 1266–1280 (2013).
- Escribano-Díaz, C. et al. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and *BRCA1*-CtIP controls DNA repair pathway choice. *Mol. Cell* **49**, 872–883 (2013).
- Xu, G. et al. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature* **521**, 541–544 (2015).
- Gupta, R. et al. DNA repair network analysis reveals shieldin as a key regulator of NHEJ and PARP inhibitor sensitivity. *Cell* **173**, 972–988 (2018).
- Dev, H. et al. Shieldin complex promotes DNA end-joining and counters homologous recombination in *BRCA1*-null cells. *Nat. Cell Biol.* **20**, 954–965 (2018).
- Noordermeer, S. M. et al. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117–121 (2018).
- Mirman, Z. et al. 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polα-dependent fill-in. *Nature* **560**, 112–116 (2018).
- Ghezraoui, H. et al. 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. *Nature* **560**, 122–127 (2018).
- Ray Chaudhuri, A. et al. Replication fork stability confers chemoresistance in *BRCA*-deficient cells. *Nature* **535**, 382–387 (2016).
- Guillemette, S. et al. Resistance to therapy in *BRCA2* mutant cells due to loss of the nucleosome remodeling factor CHD4. *Genes Dev.* **29**, 489–494 (2015).
- Rondinelli, B. et al. EZH2 promotes degradation of stalled replication forks by recruiting MUS81 through histone H3 trimethylation. *Nat. Cell Biol.* **19**, 1371–1378 (2017).
- Schlacher, K. et al. Double-strand break repair-independent role for *BRCA2* in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529–542 (2011).
- Lemaçon, D. et al. MRE11 and EXO1 nucleases degrade reversed forks and elicit MUS81-dependent fork rescue in *BRCA2*-deficient cells. *Nat. Commun.* **8**, 860 (2017).
- Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
- Planells-Cases, R. et al. Subunit composition of VRAC channels determines substrate specificity and cellular resistance to Pt-based anti-cancer drugs. *EMBO J.* **34**, 2993–3008 (2015).
- Barbar, E. Dynein light chain LC8 is a dimerization hub essential in diverse protein networks. *Biochemistry* **47**, 503–508 (2008).
- King, S. M. Dynein-independent functions of *DYNLL1*/LC8: redox state sensing and transcriptional control. *Sci. Signal.* **1**, pe51 (2008).
- The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
- Jurado, S. et al. ATM substrate Chk2-interacting Zn²⁺ finger (ASCI2) is a bi-functional transcriptional activator and feedback sensor in the regulation of dynein light chain (*DYNLL1*) expression. *J. Biol. Chem.* **287**, 3156–3164 (2012).
- Lo, K. W. et al. The 8-kDa dynein light chain binds to p53-binding protein 1 and mediates DNA damage-induced p53 nuclear accumulation. *J. Biol. Chem.* **280**, 8172–8179 (2005).
- Iacovoni, J. S. et al. High-resolution profiling of γ-H2AX around DNA double strand breaks in the mammalian genome. *EMBO J.* **29**, 1446–1457 (2010).
- Zhou, Y., Caron, P., Legube, G. & Paull, T. T. Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic Acids Res.* **42**, e19 (2014).
- Cruz-García, A., López-Saavedra, A. & Huertas, P. *BRCA1* accelerates CtIP-mediated DNA-end resection. *Cell Rep.* **9**, 451–459 (2014).
- Patch, A. M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
- Rayala, S. K. et al. Functional regulation of oestrogen receptor pathway by the dynein light chain 1. *EMBO Rep.* **6**, 538–544 (2005).
- Mazouzi, A. et al. A comprehensive analysis of the dynamic response to aphidicolin-mediated replication stress uncovers targets for ATM and ATMIN. *Cell Rep.* **15**, 893–908 (2016).
- Rapali, P. et al. Directed evolution reveals the binding motif preference of the LC8/*DYNLL* hub protein and predicts large numbers of novel binders in the human proteome. *PLoS ONE* **6**, e18818 (2011).
- Rapali, P. et al. *DYNLL*/LC8: a light chain subunit of the dynein motor complex and beyond. *FEBS J.* **278**, 2980–2996 (2011).
- Puthalakath, H., Huang, D. C., O'Reilly, L. A., King, S. M. & Strasser, A. The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex. *Mol. Cell* **3**, 287–296 (1999).
- Jaffrey, S. R. & Snyder, S. H. PIN: an associated protein inhibitor of neuronal nitric oxide synthase. *Science* **274**, 774–777 (1996).
- Dundr, M. et al. Actin-dependent intranuclear repositioning of an active gene locus in vivo. *J. Cell Biol.* **179**, 1095–1103 (2007).
- Lightcap, C. M. et al. Biochemical and structural characterization of the LC8/*DYNLL* interaction. *J. Biol. Chem.* **283**, 27314–27324 (2008).
- Song, C. et al. Serine 88 phosphorylation of the 8-kDa dynein light chain 1 is a molecular switch for its dimerization status and functions. *J. Biol. Chem.* **283**, 4004–4013 (2008).
- Jung, Y., Kim, H., Min, S. H., Rhee, S. G. & Jeong, W. Dynein light chain LC8 negatively regulates NF-κB through the redox-dependent interaction with IκBα. *J. Biol. Chem.* **283**, 23863–23871 (2008).
- Tkác, J. et al. HELB is a feedback inhibitor of DNA end resection. *Mol. Cell* **61**, 405–418 (2016).

Acknowledgements D.C. is supported by NIH grants R01 CA208244 and R01CA142698, a Leukemia and Lymphoma Society Scholar grant, and the Claudia Adams Barr Program in Innovative Basic Cancer Research. D.C. and P.A.K. are supported by DOD W81XWH-15-0564/OC140632. Y.J.H. is supported by an AACR-AstraZeneca Ovarian Cancer Research Fellowship (17-40-12-HE). J.-Y.M. was supported by a CIHR foundation grant.

Reviewer information Nature thanks T. Stracker and the anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.J.H. and D.C. designed the study with input from P.A.K. and A.D.D. Y.J.H. and K.M. performed most of the cell-based experiments with assistance from C.Y., J.B., J.M. and P.D. A.D. did the statistical analysis of images. M.-C.C., D.A.R. and J.N. conducted the in vitro studies under J.-Y.M.'s supervision. G.L. provided the guidance and reagents on the AsISI system. J.G.D. and D.E.R. provided all of the reagents and the analysis of the CRISPR library. A.S. performed statistical and computational analysis of clinical data under S.D.'s guidance. D.C. wrote the manuscript with input from P.A.K. and A.D.D.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0670-5>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0670-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to D.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Cell culture. All cells were obtained from the American Type Culture Collection (ATCC). Cells were grown in DMEM containing 10% fetal bovine serum (FBS), except KURAMOCHI cells, which were grown in RPMI-1640 (ATCC) supplemented with 10% FBS and 1% penicillin/streptomycin, and UWB1.239 cells, which were grown in RPMI-1640 supplemented with 50% RPMI-1640 and 50% mammary epithelial growth medium (MEGM) from Clonetics/Lonza, 3% FBS and 1% penicillin/streptomycin. Parental cells were tested for mycoplasma contamination.

Antibodies. Mouse antibodies used were against Flag M2, α - and β -tubulin (Sigma), BRCA1, γ -H2AX (Millipore), LIG4, GFP and c-Myc (Santa Cruz), REV7 (BD Transduction Laboratory); rabbit antibodies were against PARP1, H2AX, FEN1, RAD50 (Cell Signaling), histone H3, BLM, DNA2, DYNLL1 (Abcam), MRE11 (GeneTex), ATMIN (Millipore), EXO1 (Sigma), PTIP, 53BP1, RAD51 (Santa Cruz) and RPA32 phospho-S4/S8 (Bethyl); goat antibodies were against NBN (Santa Cruz), and rat antibody was against RPA32 (Cell Signaling).

CRISPR screen. The doses of olaparib and cisplatin used in these screens were determined by propagating UWB1.289, COV362 and JHOS-2 cells in different concentrations of these drugs to determine the effect on cell proliferation. To generate cell lines stably expressing Cas9, UWB1.289, COV362 and JHOS-2 cells were infected with the Cas9 expression vector pXPR_BRD111 and selected with $10 \mu\text{g ml}^{-1}$ blasticidin for 4–7 days. Cas9-expressing cells were maintained in 2 – $5 \mu\text{g ml}^{-1}$ blasticidin. The production of the GeCKO library from the Broad Institute (targeting 18,080 genes with 64,751 unique guide sequences) in the pLentiGuide-puro backbone has been previously described¹⁹. After transduction of library and selection for 7–10 days with puromycin, the cells were treated with olaparib or cisplatin for 14 days with medium/drug change every three days. DNA from olaparib-resistant clones and cisplatin-resistant clones were extracted with QIAGEN DNeasy Blood and tissue kits. PCR of gDNA and pDNA (sgRNA plasmid pool used to generate virus) was performed as previously described²⁰. Cells transduced with the CRISPR library but that did not go through PARPi or cisplatin selections were sequenced as input control. STARS analysis was performed using the STARS software v.1.1 (Broad Institute)²⁰. Genes were ranked based on input from all three lines and the *P* value obtained from the STARS analysis (see Supplementary Tables 1 and 2).

TCGA analysis. Publicly available TCGA ovarian serous cystadenocarcinoma data from cBioportal were queried for DYNLL1 expression data. Other clinical characteristics and BRCA1 and BRCA2 alteration data were also downloaded for the 316 ovarian carcinoma tumours in the cohort. In total, 36 carcinomas with a BRCA1 mutation, 33 BRCA1 hypermethylated tumours and 34 carcinomas harbouring a BRCA2 mutation were isolated from the cohort and used for further analysis (two tumours: TCGA-13-1512-01 and TCGA-23-1026-01 both have a BRCA1 and BRCA2 mutation and are considered to be hypermutated and not included in the analysis). All necessary calculations were performed using and Graph Pad Prism v.7.0.

Statistical analysis. All statistical calculations were performed using Excel, GraphPad Prism v.7.0 or R. Unless mentioned otherwise, all statistics were evaluated by two-tailed Student's *t*-test (Mann–Whitney *U*-test). No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 and *****P* < 0.0001.

CTG assay. To assess cellular toxicity, cells were transfected twice with siRNA at 36-h intervals. Forty-eight hours after the second transfection, cells were seeded into 96-well plates. The next day, DNA-damaging agents were added at the indicated concentrations. Cells were incubated in drug containing medium for 5 days followed by incubation in drug-free medium for 4 days before viability was measured using CellTiter-Glo reagent (Promega) and assayed using a luminescence microplate reader. Untreated and treated conditions were repeated in triplicates for each experiment, and each experiment was repeated at least three times. Survival for each drug was plotted as a percentage of survival in drug-free medium.

Immunofluorescence. Cells were grown on glass coverslips, fixed in 4% formaldehyde in PBS for 15 min at room temperature and blocked/permeabilized for 1 h in PBS containing 0.3% Triton X-100, 1% BSA, 10% FBS (or 3% goat serum). Incubation with primary and secondary antibodies (Alexa Fluor, Molecular Probes) was done in PBS containing 1% BSA and 0.1% Triton X-100 for 1 h at room temperature. Coverslips were mounted using DAPI Fluoromount-G (Southern Biotech). When staining for RPA or RAD51 foci, soluble proteins were pre-extracted as previously described⁶ before processing for immunofluorescence.

siRNA-mediated silencing. Cells were transfected with siRNAs using Lipofectamine RNAiMax following the manufacturer's instructions (Invitrogen).

The sequences of the stealth siRNAs (Thermo-Fisher) for the human genes were as follows. DYNLL1 #1: 5'-GAAGGACAUUGCGGCUCAU-3'; DYNLL1 #2: 5'-AGCCUAAAUUCCAAAUAC-3'; 53BP1: 5'-AGAACGAGGAGACGGUA

AUAGUGGG-3'; BRCA1: 5'-CAGCUACCCUCCAUAUA-3'; control: 5'-AAGCCGGUAUGCCGGUUAAGU-3'.

Protein purification. Recombinant wild-type and mutant DYNLL1, expressed in *E. coli*, and BLM, expressed in insect cells, were tagged at the N terminus with GST and at the C terminus with His10 using a protocol described previously⁴². MRE11, RAD50 and NBS1 were purified according to an established protocol⁴³. EXO1 was purified according to a previous study⁴¹ and MRE11 was purified as previously described⁴⁴. RPA was purified as described⁴⁵. For recombinant DNA2 protein purification, SF9 insect cells (1×10^6 cells ml^{-1}) were infected with GST–DNA2–Flag baculovirus. At 48 h post-infection, cells were collected by centrifugation and the pellet was frozen on dry ice. Cells were lysed in buffer 1 ($1 \times$ PBS containing 150 mM NaCl, 1 mM EDTA, 0.05% Triton X-100, 1 mM DTT and protease inhibitors) and homogenized by 20 passes through a Dounce homogenizer (pestle A). Cell lysate was incubated with 1 mM MgCl_2 and 2.5 U ml^{-1} benzonase nuclease at 4°C for 1 h followed by centrifugation at $90,000g$ for 1 h. Soluble cell lysate was incubated with 1 ml of GST–Sepharose beads for 90 min at 4°C with gentle rotation. Beads were washed twice with buffer 1 followed by incubation with buffer 2 (buffer 1 with 5 mM ATP and 15 mM MgCl_2) for 1 h at 4°C . Sepharose–GST beads were washed twice with buffer 1 supplemented with 200 mM NaCl and once with P5 buffer (50 mM NaH_2PO_4 pH 7.0, 500 mM NaCl, 10% glycerol, 0.05% Triton X-100 and 5 mM imidazole) followed by cleavage with PreScission protease (60 U ml^{-1} , GE Healthcare Life Sciences), overnight at 4°C in P5 buffer. Supernatant was then collected and completed to 10 ml with Flag binding buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 10% glycerol and 0.025% Triton X-100) before to incubate with $600 \mu\text{l}$ of M2 anti-Flag affinity gel (Invitrogen) for 1 h at 4°C . Beads were washed twice with washing buffer (Flag binding buffer supplemented with 100 mM NaCl). After two additional washes with elution flag buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.025% Triton X-100 and 10% glycerol), proteins were eluted twice in one volume of beads with elution flag buffer and $500 \mu\text{g ml}^{-1}$ of 3 \times -Flag peptide for 45 min at 4°C . Protein was then dialysed in the storage buffer (20 mM Tris–HCl, pH 7.4, 200 mM NaCl, 10% glycerol and 1 mM DTT) and stored in aliquots at -80°C .

In vitro resection assay (BLM–DNA2). In vitro resection experiments were performed using pUC19 DNA linearized with KpnI and then 3' labelled with [α - ^{32}P]ATP and terminal deoxytransferase (NEB). Reactions were conducted using 50 nM of linearized DNA substrate in standard buffer (20 mM Na-HEPES pH 7.5, 0.05% Triton X-100, $100 \mu\text{g ml}^{-1}$ BSA). Buffer was supplemented with 2 mM ATP and 5 mM MgCl_2 immediately before addition of the purified proteins. Reactions were initiated on ice by adding wild-type DYNLL1 or the mutant versions and transferred immediately to 37°C . After 5 min, the order of addition and incubation of the respective protein components were: MRN (10 nM, 5 min), RPA (100 nM, 4 min), BLM (6.5 nM, 4 min) and DNA2 (10 nM, 45 min). Reactions were followed by proteinase K treatment for 1 h. Products were analysed on a 1% native agarose gel. Gels were dried on DE81 paper (Whatman) and signals were detected by autoradiography. Densitometric analyses were performed using the FLA-5100 phosphorimager (Fujifilm) and quantitated using the Image Reader FLA-5000 v1.0 software.

In vitro resection assay (MRN). The dsDNA probe for the MRN gel was made by annealing the following two oligonucleotides followed by purification, as described previously⁴⁶. Forward: GGGCGAATTGGGCCGACGTCGCA TGCTCCTCTAGACTCGAGGAATTCGGTACCCCGGGTTCGAAATCGATA AGCTTACAGTCTCCATTTAAAGGACAAG; reverse: CTTGTCTTTAAATG GAGACTGTAAGCTTATCGATTTCGAACCCGGGGTACCGAATTCCTCGA GTCTAGAGGAGCATGCGACGTCGGGCCCAATTCGCCC.

SMART assay. DNA fibre analysis was performed as previously described⁴⁷. In brief, cells were transfected with siRNA for 48 h with bromodeoxyuridine (BrdU) in culture medium to label all genomic DNA. After 48-h treatment with $10 \mu\text{M}$ olaparib, cells were spotted on a slide followed by lysis with lysis buffer (200 mM Tris–HCl pH 5.5, 50 mM EDTA and 0.5% SDS), the denature step for normal BrdU staining was skipped so that BrdU antibody will only recognize ssDNA labelled with BrdU. After three washes in water, slides were processed for immunofluorescence. Coverslips were blocked for 1 h in PBS containing 1% BSA. Incubation with primary antibodies rat anti-BrdU (BD Biosciences) and mouse anti-dsDNA antibody (Millipore) was done in PBS containing 1% BSA for 1 h at 37°C . Coverslips were washed three times in PBS followed by incubation with secondary antibodies goat anti-rat IgG Alexa Fluor 488 and donkey anti-mouse IgG Alexa Fluor 594 (both Life Technologies) for 1 h at 37°C . Coverslips were mounted on slides with DAPI Fluoromount-G (Southern Biotech). Images were analysed using ImageJ software.

RNA purification and qPCR. RNA was purified using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. cDNA was generated from $1 \mu\text{g}$ of purified RNA using a Quanta Biosciences qScript cDNA Synthesis Kit according to the manufacturer's protocol. One microlitre out of $20 \mu\text{l}$ cDNA was used to perform qPCR using the NEB supermix with the following gene-specific

primers: *BLM* forward #1 5'-GTGTTACACCACCCAAAGTC-3', *BLM* reverse #1 5'-GGAGGCAAATCAGTCTTTACTG-3'; *BLM* forward #2 5'-5GGACCTTGACACCTCTGACAG-3', *BLM* reverse #2 5'-GGATTCAGCTCCTGCATACTC-3'; *DNA2* forward #1 5'-GTGCCATACCTGTCACAAATC-3', *DNA2* reverse #1 5'-GAAGGACCGACAAGTTTCTGTC-3'; *DNA2* forward #2 5'-CAGAACTTGTCGGTCTTCC-3', *DNA2* reverse #2 5'-GTGGAAGAACAGAAGTAAAGTAGG-3'; *EXO1* forward #1 5'-CTAGCCAAAGGTGAACCTACTG-3', *EXO1* reverse #1 5'-GTGTGATATTGATAGACCGGGTG-3'; *EXO1* forward #2 5'-CCTCGTGGCTCCCTATGAAG-3', *EXO1* reverse #2 5'-CTAGGAGATCCGAGTCCTCTG-3'; *NBN* forward #1 5'-TGGAGCAGGAAAACCTCCAC-3', *NBN* reverse #1 5'-GATTTCTGCCTTAGCCACT-3'; *NBN* forward #2 5'-CACTCACCTTGTCATGGTATCAG-3', *NBN* reverse #2 5'-CTGCTTCTTGGACTCAACTGC-3'; *MRE11A* forward #1 5'-CTTGACGACTGCGAGTGA-3', *MRE11A* reverse #1 5'-TTCACCCATCCCTCTTTCTG-3'; *MRE11A* forward #2 5'-GCTCTTCTCTTTGAGACCC-3', *MRE11A* reverse #2 5'-TCTGCCTTTA GTGCTGATGAC-3'; *RAD50* forward #1 5'-CAGGAGGGAATCTCCAGTCAA-3', *RAD50* reverse #1 5'-TTTGGTTGGACCCAATGGGG-3'; *RAD50* forward #2 5'-TACTGGAGATTTCCTCCTGG-3', *RAD50* reverse #2 5'-AGACTGACCTTTTACCATGC-3'.

Genomic and transcriptomic analysis. Genomic and transcriptomic data for chemo-resistant, HGSC was obtained from the Australian Ovarian Cancer Study cohort (OV-AU), and Pancreatic Cancer Endocrine neoplasms (PAEN-AU) cohorts, which are part of the International Cancer Genome Consortium (ICGC). Inactivating germline or somatic mutations in genes associated with homologous recombination repair, or BRCA1 methylation, were detected collectively in half of the primary tumours. For these two cohorts, somatic structural variants (duplications, deletions, inversions, intra- and interchromosomal translocations) were identified using qSV (PMCID: PMC4523082) from the whole-genome sequencing data for 93 and 33 patients, respectively. There were 48–2,431 (median: 292) structural variants per sample in the ovarian cohort and 4–154 (median: 18) structural variants per sample in the pancreatic cohort. RNA-sequencing-based expression data were also available for these tumours from the ICGC data portal (<https://dcc.icgc.org/>). We grouped the samples into four categories based on combinatorial high (above median) and low (below median) expression of BRCA1 and DYNLL1; this enabled us to analyse samples with BRCA deficiency due to classic BRCA1 inactivation or loss as well as those potentially mediated by other mechanisms (for example, promoter hypermethylation). Statistical analysis was performed using R.

Plasmids and transfection. Two CRISPR guide RNAs were selected from GeCKO library. sgRNAs targeting the *ATMIN* and *DYNLL1* loci were cloned in pLentiGuide-puro vector (Addgene 52963). Then 24 h after transduction, cells

were selected with puromycin. Unless otherwise mentioned, stable and transient transfections were performed using Lipofectamine 2000 (Invitrogen) or Eugene 6 (Promega) following the manufacturer's instructions.

Immunoprecipitation. Unless otherwise mentioned, proteins were immunoprecipitated from whole-cell extracts. In brief, cells were collected, washed and lysed for 30 min in a buffer containing 20 mM Tris-HCl (pH 7.65), 300 mM NaCl, 0.5% NP-40, 5 mM EDTA, 5% glycerol and protease and phosphatase cocktail inhibitors (Roche). Protein concentration from cleared supernatants was estimated using Bradford assay (Biorad). Whole-cell extracts (2 mg) were incubated on a roller for 16 h at 4°C with anti-Flag (Sigma). Resins were washed five times with TGN buffer (20 mM Tris-HCl pH 7.65, 150 mM NaCl, 3 mM MgCl₂, 10% glycerol and 0.5% NP-40). Eluted proteins were analysed by immunoblotting.

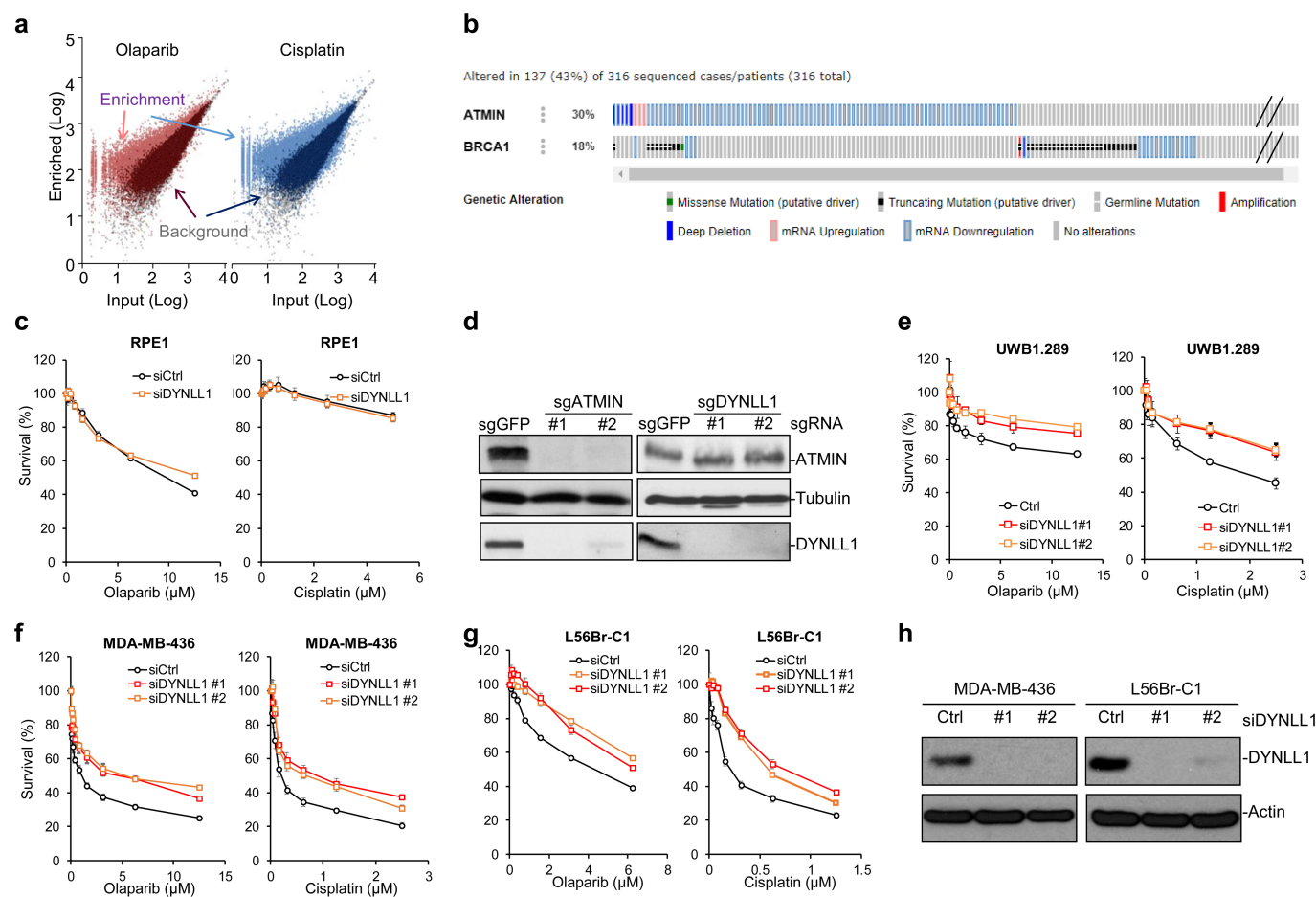
GST pull-down. GST, GST-DYNLL1(WT) or GST-DYNLL1(Ser88Asp) bound on Glutathione Sepharose beads (GE Healthcare) was incubated in 500 µl GSTB buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% NP40, 5 mM NaF, 1 mM Na₃VO₄, protease inhibitor cocktail (Roche) and 1 mg ml⁻¹ BSA) for 20 min at room temperature. Purified MRE11 protein, or the indicated proteins, were then added to each reaction and incubated for 20 min at room temperature. Complexes were washed four times with GSTB buffer without BSA. Proteins were visualized by western blotting using the indicated antibodies.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

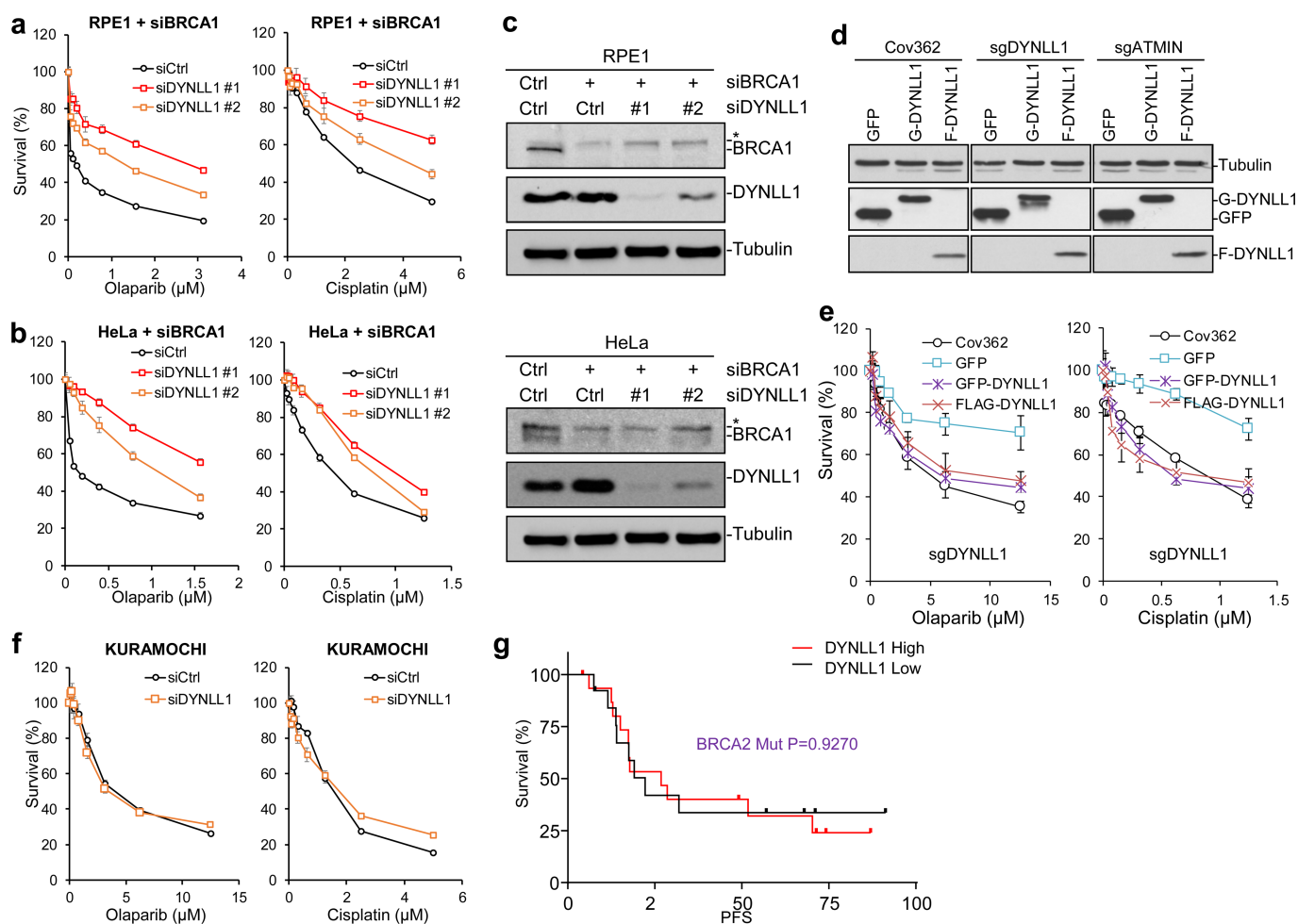
All relevant data are included in the paper and/or its Supplementary Information.

42. Buisson, R. et al. Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. *Nat. Struct. Mol. Biol.* **17**, 1247–1254 (2010).
43. Yu, Z. et al. The MRE11 GAR motif regulates DNA double-strand break processing and ATR activation. *Cell Res.* **22**, 305–320 (2012).
44. Boisvert, F.-M., Déry, U., Masson, J.-Y. & Richard, S. Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. *Genes Dev.* **19**, 671–676 (2005).
45. Henriksen, L. A., Umbricht, C. B. & Wold, M. S. Recombinant replication protein A expression, complex formation, and functional characterization. *J. Biol. Chem.* **269**, 11121–11132 (1994).
46. Moiani, D. et al. Targeting allostery with avatars to design inhibitors assessed by cell activity: dissecting MRE11 endo- and exonuclease activities. *Methods Enzymol.* **601**, 205–241 (2018).
47. Nieminiusz, J., Schwab, R. A. & Niedzwiedz, W. The DNA fibre technique - tracking helicases at work. *Methods* **108**, 92–98 (2016).



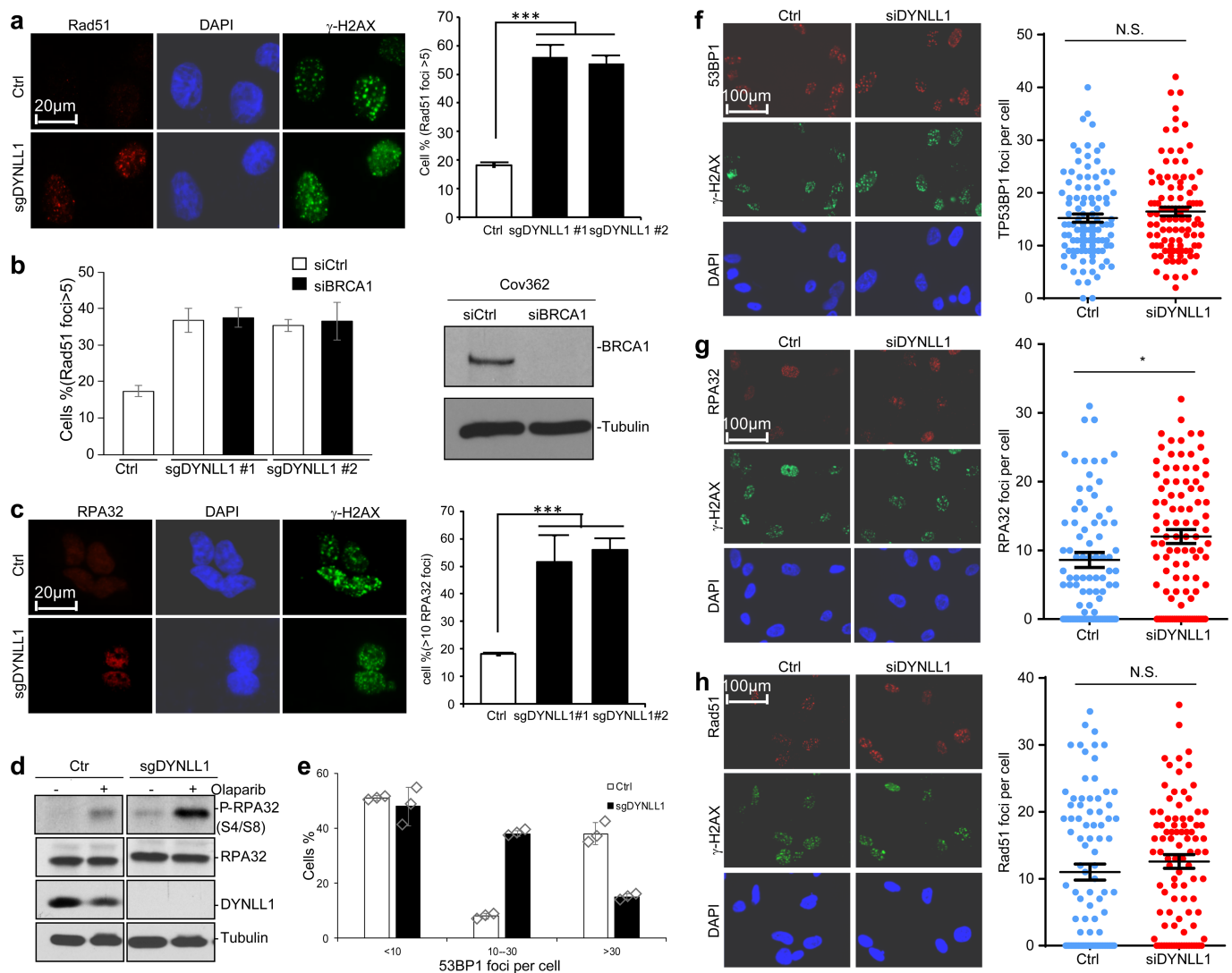
Extended Data Fig. 1 | DYNLL1 depletion causes resistance to PARPi and cisplatin in multiple lineages. a, Relative guide abundance before and after olaparib and cisplatin treatment in Cov362 cells (data provided in Supplementary Tables 1 and 2). **b**, Comparison of ATMIN and BRCA1 alterations in ovarian cancer according to the TCGA dataset²⁴ (316 samples) from the cBioPortal. **c**, Survival assay of RPE1 cells treated with olaparib (left) or cisplatin (right), after transfection with non-targeting control or *DYNLL1* siRNA (siCtrl or siDYNLL1). **d**, Immunoblot of

ATMIN and DYNLL1 from Cov362 cells with deletions of ATMIN or DYNLL1 (sgATMIN or sgDYNLL1). Tubulin was used as a loading control. **e–g**, Survival assay of *BRCA1*-mutant cells UWB1.289 (**e**), MDA-MB-236 (**f**) and L56Br-C1 (**g**) treated with olaparib or cisplatin, and transfected with control or *DYNLL1* siRNA. Data are mean \pm s.e.m. from three different experiments. **h**, Immunoblots showing depletion of DYNLL1. Experiments were repeated independently three times with similar results.



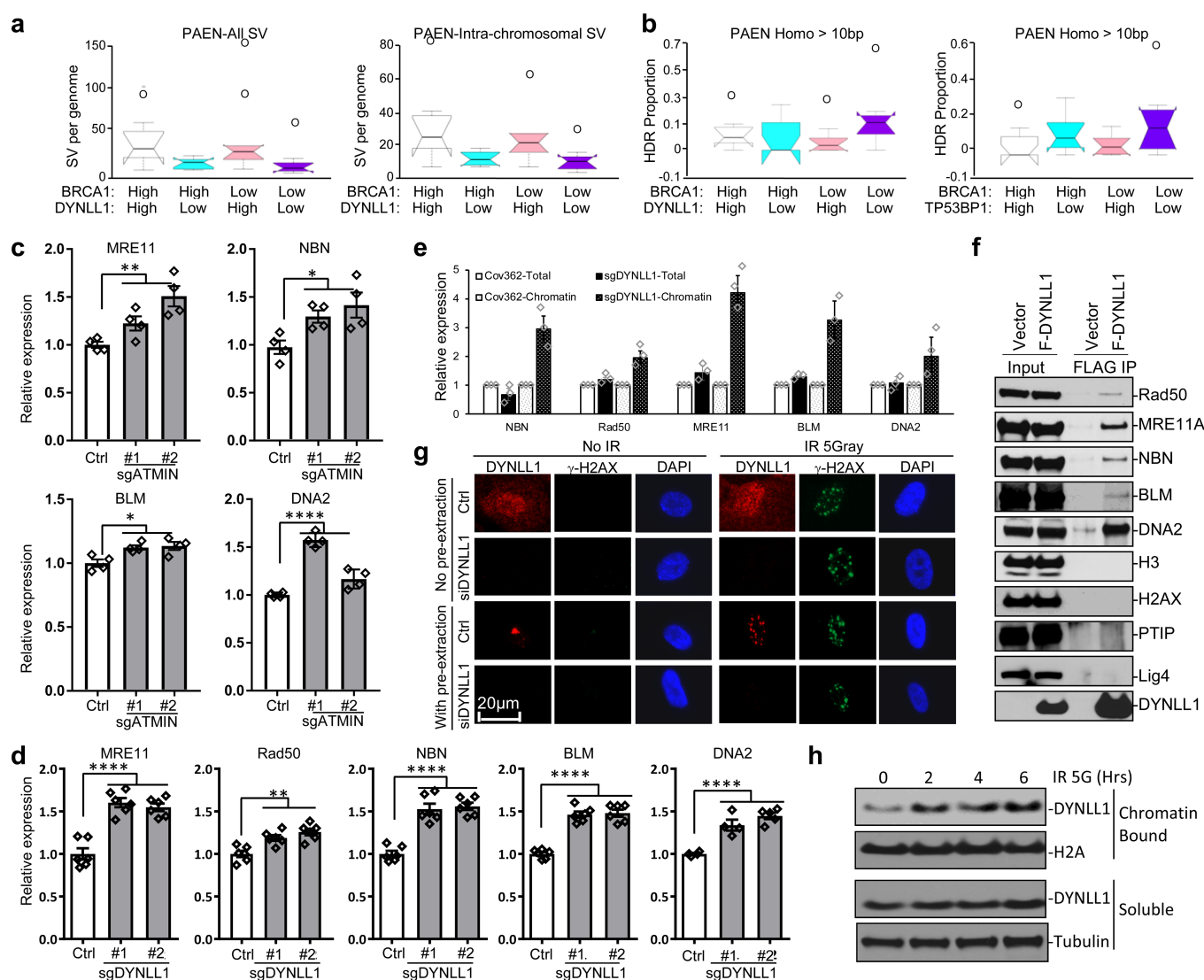
Extended Data Fig. 2 | Impact of DYNLL1 on PARPi and cisplatin is specific to BRCA1-mutant cells. **a, b**, Survival assay of RPE1 (**a**) and HeLa (**b**) cells transfected with *BRCA1* siRNA and treated with olaparib (left) or cisplatin (right), and co-transfected with control or *DYNLL1* siRNA. **c**, Immunoblots showing depletion of DYNLL1 and *BRCA1*. Experiments were repeated independently three times with similar results. #1 and #2 are independent stable clones. **d**, Immunoblot of tagged DYNLL1 (G, GFP; F, Flag) in Cov362 cells after deletions of *ATMIN* or *DYNLL1* (sgATMIN or sgDYNLL1). **e**, Survival assay of Cov362 *DYNLL1*^{-/-} clone expressing tagged DYNLL1, treated with olaparib

(left) or cisplatin (right). Data are mean \pm s.e.m. from three different experiments. **f**, Survival assay of the indicated Cov362 clones transfected with KURAMOCHI cells and control or *DYNLL1* siRNA, and treated with olaparib or cisplatin. For all panels, data are mean \pm s.e.m. from three different experiments. **g**, PFS of ovarian carcinoma patients with *BRCA2* mutation based on above or below median expression values of DYNLL1 (DYNLL1-high $n = 14$, DYNLL1-low $n = 18$; source: ovarian cancer, TCGA dataset²⁴). Statistical significance was assessed by the one-sided Mantel-Cox test.



Extended Data Fig. 3 | DYNLL1 influences RAD51 foci and RPA32 foci formation in BRCA1-mutant cells. a, b, Immunofluorescence and quantification of RAD51 foci (**a, b**) and RPA32 foci (**c**) in wild-type and *DYNLL1*^{-/-} Cov362 cells exposed to 5 Gy ionizing radiation. Staining is 6 h (RAD51) and 4 h (RPA32) after ionizing radiation. In **a**, $n = 105$; *** $P < 0.0001$ (control versus sgDYNLL1 #1), *** $P = 0.0003$ (control versus sgDYNLL1 #2), $P = 0.5679$ (sgDYNLL1 #1 versus sgDYNLL1 #2); two-tailed unpaired Student's *t*-test. In **b**, wild-type and *DYNLL1*^{-/-} Cov362 cells were also transfected with control and *BRCA1* siRNA and immunoblotting was used to confirm silencing. Data are mean \pm s.e.m. from three different experiments ($n = 100$). In **c**, $n = 100$; *** $P = 0.0002$ (control versus sgDYNLL1 #1), *** $P < 0.0001$ (control versus sgDYNLL1 #2), $P = 0.5679$ (sgDYNLL1 #1 versus sgDYNLL1 #2); unpaired two-tailed

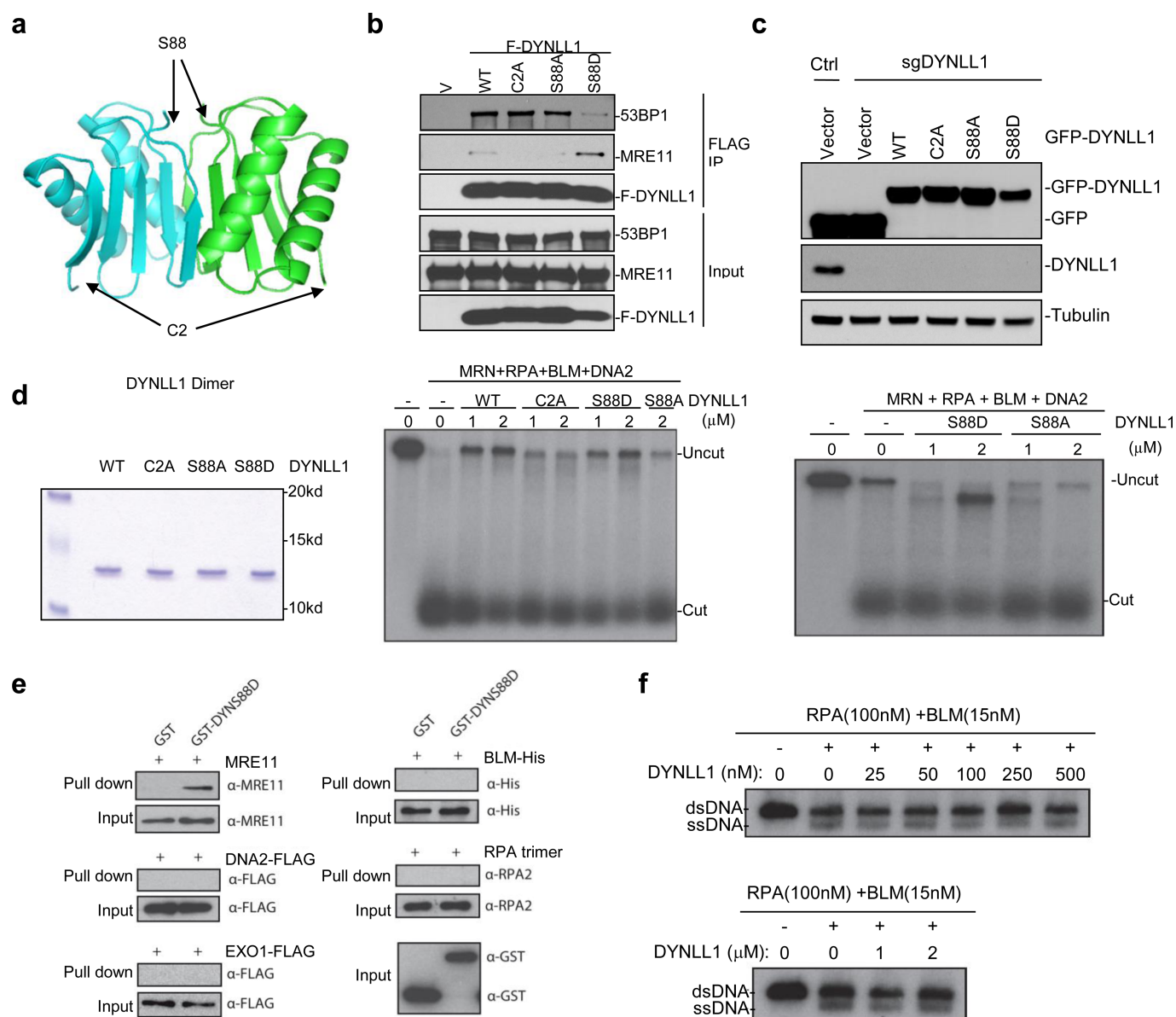
Student's *t*-test. **d**, Immunoblot of wild-type and *DYNLL1*^{-/-} Cov362 cells after 10 μ M olaparib treatment for 48 h with indicated antibodies. Experiments were repeated independently for three times with similar results. **e**, Analysis of 53BP1 foci as shown in Fig. 2b from wild-type and *DYNLL1*^{-/-} Cov362 cells treated with 10 μ M olaparib for 24 h. Data are mean \pm s.e.m. from three different experiments ($n = 102$ cells of each genotype). **f-h**, Immunofluorescence (left) and quantification (right) of 53BP1 foci (**f**; $n = 95$; $P = 0.1019$), RPA32 foci (**g**; $n = 100$; * $P = 0.0238$) and RAD51 foci (**h**; $n = 94$, $P = 0.3161$) in RPE1 cells transfected with control or *DYNLL1* siRNA exposed to 5 Gy ionizing radiation for 1 h (53BP1, **f**), 4 h (RPA32, **g**) or 6 h (RAD51, **h**). Statistical analyses were by unpaired two-sided Student's *t*-test. For all panels, data are mean \pm s.e.m. from three different experiments.



Extended Data Fig. 4 | DYNLL1 regulates the DNA end-resection

machinery. **a**, Samples from the PAEN-AU cohort were grouped into four categories based on combinatorial high (above median) and low (below median) expression levels of BRCA1 and DYNLL1. The frequency of somatic structural variants (deletion, duplication, insertion or intrachromosomal translocation) (left) and the frequency of intrachromosomal structural variants (deletion, duplication, insertion or intrachromosomal translocation) (right) were plotted. **b**, Samples from the PAEN-AU (32 samples) cohort were grouped into four categories based on combinatorial high (above median) and low (below median) expression levels of BRCA1 and DYNLL1 (left) or of BRCA1 and 53BP1 (right). The frequency of structural variants (deletions, duplication, insertions or intrachromosomal translocation) with indications of homology-directed repair (≥ 10 -bp homology) was plotted. In all box plots, the upper whisker is $1.5 \times$ IQR more than the third quartile, and the lower whisker is $1.5 \times$ IQR lower than the first quartile, respectively, in which the interquartile range (IQR) is the difference between the third and the first quartile (that is, the box length). Circles denote outliers. **c**, Quantification of mRNA levels of the indicated genes in control and *ATMIN*^{-/-} (sgATMIN) cells ($n = 4$). Expression levels were normalized to *ACTB*. Data are mean \pm s.e.m. from four different experiments. * P = 0.0335 and ** P = 0.0038 (control versus #1 and #2, respectively, for *MRE11*); * P = 0.0152 and * P = 0.0257 (control versus #1 and #2, respectively, for *NBN*); * P = 0.0130 and * P = 0.0203 (control versus #1 and #2, respectively, for *BLM*), **** P < 0.0001 and * P = 0.0179 (control versus #1 and #2, respectively, for *DNA2*);

unpaired two-sided Student's t -test. **d**, Quantification of mRNA levels of the indicated genes in control and *DYNLL1*^{-/-} (sgDYNLL1) Cov362 cells ($n = 6$). Expression levels were normalized to *ACTB*. Data are mean \pm s.e.m. from six different experiments. **** P < 0.0001 and **** P < 0.0001 (control versus #1 and #2, respectively, for *MRE11*); **** P < 0.0001 (control versus #1 and #2, respectively, for *RAD50*); **** P < 0.0001 and **** P < 0.0001 (control versus #1 and #2, respectively, for *NBN*); **** P < 0.0001 and **** P < 0.0001 (control versus #1 and #2, respectively, for *BLM*); **** P = 0.0002 and **** P < 0.0001 (control versus #1 and #2, respectively, for *DNA2*); unpaired two-sided Student's t -test. **e**, Quantification of subcellular fraction of indicated proteins ($n = 3$) in control and *DYNLL1*^{-/-} Cov362 cells. Levels of total protein and chromatin-bound protein were normalized to H2AX levels, and levels of indicated proteins in *DYNLL1*^{-/-} Cov362 cells are graphically represented relative to the control Cov362 cells. Data are mean \pm s.e.m. **f**, Flag immunoprecipitation of Flag-DYNLL1 and immunoblot with indicated antibodies. **g**, Immunofluorescence and quantification of DYNLL1 and γ -H2AX foci in RPE1 cells transfected with control and *DYNLL1* siRNA, 1 h after 5 Gy ionizing radiation (IR). Experiments were repeated independently three times with similar results. **h**, Immunoblot of DYNLL1 in RPE1 cells exposed to 5 Gy ionizing radiation and subcellular fractionation at indicated times. Experiments were repeated independently three times with similar results.



Extended Data Fig. 5 | Separation of the functions of DYNLL1 mutants that influence DNA end resection in vitro. **a**, Structure of DYNLL1 dimer with potentially relevant residues indicated. **b**, Immunoprecipitation of indicated DYNLL1 mutants with 53BP1 and MRE11. Experiments were repeated independently three times with similar results. **c**, Immunoblot of tagged wild-type and mutant DYNLL1 in *DYNLL1*^{-/-} Cov362 cells from Fig. 4b. **d**, Resection products of wild-type or mutant recombinant DYNLL1 (purified proteins, left panel) with MRN-RPA-BLM-DNA2 and a ³²P-labelled linear 2.7-kb dsDNA substrate. Experiments were

repeated independently three times with similar results. **e**, GST pull-down of GST-tagged mutant DYNLL1 (Ser88Asp) incubated with purified human MRE11 or human DNA2, EXO1, BLM or the human RPA trimer (RPA70-RPA32-RPA14). Experiments were repeated independently three times with similar results. **f**, Recombinant wild-type DYNLL1 protein was incubated with RPA and BLM and with a ³²P-labelled linear 2.7-kb dsDNA substrate to monitor DNA unwinding. Experiments were repeated independently three times with similar results.

Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register

V. Negnevitsky^{1,2*}, M. Marinelli^{1,2*}, K. K. Mehta¹, H.-Y. Lo¹, C. Flühmann¹ & J. P. Home^{1*}

Quantum error correction is essential for realizing the full potential of large-scale quantum information processing devices^{1,2}. Fundamental to its experimental realization is the repetitive detection of errors via projective measurements of quantum correlations among qubits, as well as corrections using conditional feedback³. Repetitive application of such tasks requires that they neither induce unwanted crosstalk nor impede further control operations, which is challenging owing to the need to dissipatively couple qubits to the classical world for detection and reinitialization. For trapped ions, state readout involves scattering large numbers of resonant photons, which increases the probability of stray light causing errors on nearby qubits and leads to undesirable recoil heating of the ion motion. Here we demonstrate up to 50 sequential measurements of correlations between two beryllium ion microwave qubits using an ancillary optical qubit in a calcium ion, and implement feedback that allows us to stabilize two-qubit subspaces as well as Bell states, a class of maximally entangled states. Multi-qubit mixed-species gates are used to transfer information within the register from the qubit to the ancilla, enabling readout with negligible crosstalk to the data qubits. Heating of the ion motion during detection is mitigated by recooling all three ions using light that interacts with only the calcium ion, known as sympathetic cooling. A key element of our experimental setup is a powerful classical control system that features flexible in-sequence processing for feedback control. The methods employed here provide essential tools for scaling trapped-ion quantum computing, quantum-state control and entanglement-enhanced quantum metrology⁴.

Correction of errors arising from noise and imperfect gate operations represents a primary challenge in the development of quantum computers⁵. Quantum error correction (QEC) involves encoding the states of one logical qubit into codewords on multiple physical qubits. Repeated measurement of multi-qubit correlations in a subset of the physical qubits in the code allows errors to be detected without disturbing the stored information. To reverse the detrimental effect of errors, information of the measurement result must be processed in real time so that appropriate correction operations can be chosen and applied. Such feedback allows the quantum system to be stabilized throughout the extended periods required for computation⁶.

QEC codes have been demonstrated in several systems^{7–9}. Many early demonstrations relied on protocols that brought qubits out of the code space for error syndrome measurement. For practical QEC on continuously encoded information, ideal von Neumann projective measurements of multi-qubit operators are required³. These have been realized with trapped ions¹⁰ and in solid-state systems¹¹, including experiments in which up to three rounds of feedback were conditioned on the result of the measurement^{12,13}. However, to perform useful repeated quantum measurement and feedback¹⁴, as required for indefinite stabilization of quantum systems, additional stringent conditions must be satisfied: (i) the measurement time must be short compared to the relevant decoherence times of the system, (ii) the measurement process should not adversely affect the stored quantum information,

and (iii) the measurement process should not impede our ability to perform subsequent measurements. Ancilla measurement and reset fundamentally require dissipative operations that couple the ancilla to the classical world, and preserving the quantum state of the data qubits places strong requirements on suppression of crosstalk during such operations. Although up to three consecutive measurements have been performed on a multi-qubit trapped-ion register by hiding data qubits in internal states that do not interact with the resonant light used for detection¹⁵, the implementation of this technique with high fidelity in large-scale systems is challenging. An alternative approach is offered by the use of two species, one for the ancilla and one for the data qubits, which ensures a high degree of spectral isolation^{13,16,17}. This also provides the possibility to mitigate errors due to ion heating and transport using sympathetic cooling^{18,19}. Non-destructive sampling of quantum information from multiple qubits has application beyond quantum computing, particularly in the field of metrology, where it provides direct access to the correlation functions of the system²⁰.

Here, we demonstrate the repeated readout of two-qubit correlations stored in the hyperfine ground-state structure of two beryllium ions (⁹Be⁺) by using an ancilla qubit stored in an optical transition in calcium (⁴⁰Ca⁺). Single- and multi-qubit operations are performed on the three trapped ions using Raman laser fields for Be⁺ and a narrow-line-width laser for Ca⁺ (see Methods). We achieve long beryllium qubit coherence times compared to the time required for all operations by encoding quantum information on a field-insensitive transition²¹ in Be⁺, while the use of a different species (calcium) as the ancilla minimizes crosstalk from the measurement^{22,23}. Ancilla recycling is performed by optical pumping of the calcium ion after each measurement, and a combination of electromagnetically induced transparency (EIT)²⁴ and sideband cooling is used to re-initialize the motional ‘quantum bus’ that is used for the quantum gates between the Be⁺ and Ca⁺ qubits. These tools allow repeated measurements for more than 50 cycles. In addition, we apply feedback operations involving dynamic updates to both the beryllium–laser sequences and the trapping potentials using a control system with an embedded central processing unit, which computes the appropriate operations during the experimental sequence. Sequential measurements in different bases allow preparation of Bell states from arbitrary input states, with feedback allowing entanglement to be stabilized over extended periods.

We measure the commuting stabilizer operators $S_Z = Z_1 \otimes Z_2$ and $S_X = X_1 \otimes X_2$ (indices refer to the qubit number and X, Y, Z are the Pauli operators) using the circuit shown in Fig. 1a¹⁰. Eigenstates of these operators with eigenvalues $E_{Z/X} = +1$ (-1) have positive (negative) parity. To map the value of the parity onto the calcium measurement basis, we apply the multi-qubit operation U_{S_Z} to the beryllium qubits with the calcium ion initialized in $|0\rangle$. A subsequent projective measurement of calcium using state-dependent fluorescence then completes the stabilizer readout M_{S_Z} . For an ideal implementation, this projects Be⁺ into a $+1$ (-1) eigenstate of S_Z correlated with the Ca⁺ ion detected in $|1\rangle$ ($|0\rangle$). At the core of U_{S_Z} is a diagonal operation in the computational basis, which can be written as $\exp(i\pi Z_{Ca} \otimes S_Z/4)$ (blue boxes in

¹Institute for Quantum Electronics, ETH Zürich, Zürich, Switzerland. ²These authors contributed equally: V. Negnevitsky, M. Marinelli. *e-mail: nvlad@phys.ethz.ch; mmatteo@phys.ethz.ch; jhome@phys.ethz.ch

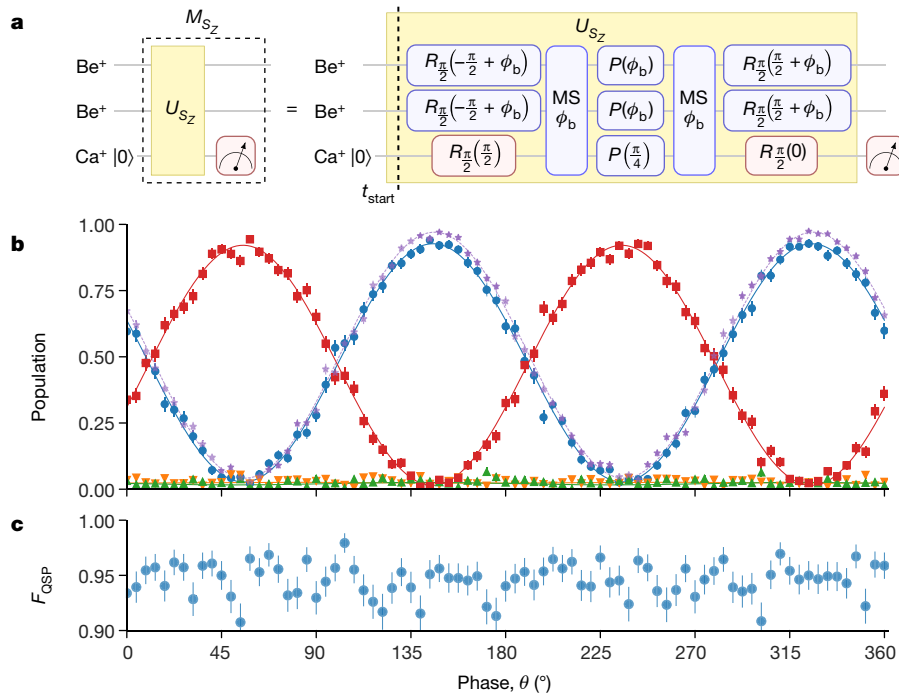


Fig. 1 | S_z parity measurement. **a**, The parity measurement M_{S_z} involves a unitary U_{S_z} , which entangles the Be^+ parity eigenspaces with the Ca^+ state, followed by a projective measurement of the latter. The decomposition of U_{S_z} , shown on the right, includes two MS gates, $R_{\pi/2}(\phi)$ -basis rotations and inversion operations $P(\phi) \equiv \cos(\phi)X + \sin(\phi)Y$. The dashed vertical line indicates the reference time t_{start} for the local phase accumulation (see Methods) and ϕ_b represents the Raman beam difference phase, to which the sequence is designed to be insensitive (see main text). The measurement basis can be rotated to S_X using $R_{\pi/2}$ pulses on the beryllium ions before and after U_{S_z} (not shown). **b**, Data from the calcium detection of a single round of parity measurements combined with the direct detection of beryllium after the measurement as a function of the phase θ for the two-ion Be^+ input state $\cos(2\theta)|-\rangle + \sin(2\theta)|+\rangle$, where $|+\rangle = |00\rangle + e^{i\phi(\theta)}|11\rangle$, $E_z = +1$ and $|-\rangle = |01\rangle + |10\rangle$, $E_z = -1$. θ is offset

Fig. 1a) and which is implemented using two multi-species three-qubit Mølmer–Sørensen (MS) gates²⁵ (applied using the in-phase motional mode) plus additional single-qubit rotations¹⁰. To map the phase shift of the calcium ion onto the measurement basis, we embed this operation between $\pi/2$ rotations (defined as $R_{\pi/2}(\phi) \equiv \exp[i\pi P(\phi)/2]$ with $P(\phi) = \cos(\phi)X + \sin(\phi)Y$) that differ in phase by $\pi/2$, where ϕ is the phase. This can be interpreted as a Ramsey interference experiment. The resulting unitary is $U_{S_z} = R_{\pi/2}^{\text{Ca}}(\pi/2)\exp(i\pi Z_{\text{Ca}} \otimes S_z)R_{\pi/2}^{\text{Ca}}(0)$.

Readout of S_X is performed by a unitary U_{S_X} , which is formed by embedding U_{S_z} between $R_{\pi/2}$ -basis rotations of both beryllium qubits about the Y axis. These are performed using Raman transitions driven by two co-propagating laser beams. Although for an ideal implementation of S_X it is theoretically possible to use a reduced pulse sequence, we operate under the experimental constraint that the parity readout must perform the same operation at any point in the sequence when it is applied. The MS gate between the hyperfine ground-state qubits of the two Be^+ ions and the optical Ca^+ qubit uses a pair of perpendicular beams for the Be^+ qubit motion coupling, whose relative phase drifts over many measurement rounds. The gate produces the operation $\text{MS} \equiv \exp(i\pi \Pi_{\phi_b}^2/8)$ with $\Pi_{\phi_b} = X_{\text{Ca}} + P_1(\phi_b) + P_2(\phi_b)$, where ϕ_b is directly proportional to the phase difference of the Raman beams. Because this difference varies little on the 160- μs timescale of a single application of the unitary, we are able to make the unitary insensitive to ϕ_b by using the same Raman beam pair to perform $\pi/2$ qubit rotations before the first and after the second MS gate in the unitary. This results in a unitary diagonal in the computational basis and thus ensures insensitivity to ϕ_b ^{23,26}. This choice also allows us to simplify pulse sequence control (see Methods).

from the phase ϕ_p (see main text) owing to an uncompensated Stark shift in the state preparation. Blue circles (red squares) show the probability of observing both the Be^+ in the $E_z = +1$ ($E_z = -1$) subspace and measuring the Ca^+ to be in $|1\rangle$ ($|0\rangle$). Orange (green) triangles indicate the undesired populations for which the Be^+ is in $E_z = +1$ ($E_z = -1$) and Ca^+ is in $|0\rangle$ ($|1\rangle$)—that is, the results are anti-correlated with respect to the ideal case. Violet stars show the parity of the input state, expressed as $(\langle S_z \rangle + 1)/2$, which is measured using direct beryllium detection in a separate experiment, where M_{S_z} is not implemented. **c**, The corresponding quantum state fidelity, which indicates the mean conditional probability that the parity subspace in the beryllium detection is correct, given the value obtained from M_{S_z} . Each point is the result of 300 experiments and error bars are 1σ uncertainties from quantum projection noise.

To test the performance of a single round of the parity measurement M_{S_z} , we input states with a range of parities and compare the Ca^+ detection results to the results of a subsequent state detection performed directly on the Be^+ ions. To prepare the input states, we create the Bell state $(|00\rangle - i|11\rangle)/\sqrt{2}$ using an MS gate acting only on the two Be^+ ions, and then we apply an $R_{\pi/2}(\phi_p)$ pulse to both ions with phase ϕ_p . For $\phi_p = 3\pi/4$, the resulting state has $E_z = -1$, whereas for $\phi_p = \pi/4$, $E_z = +1$. Results of both the Ca^+ and Be^+ measurements are shown in Fig. 1b. For comparison, we also plot the parity obtained from measuring the input state of the Be^+ ions in a separate reference experiment. We can compare the probability distributions estimated from the Be^+ measurements with and without the M_{S_z} parity measurement using a classical fidelity, for which the average value across all input states is the non-demolition fidelity $F_{\text{ND}} = 98.4(2)\%$. A more important quality measure for the QEC is the quantum state preparation fidelity F_{QSP} , the conditional probability for a given detection result to project the system into the correct subspace. As shown in Fig. 1c, this is largely independent of the input state and has a mean value of $F_{\text{QSP}} = 94.5(3)\%$. These fidelities are mathematically defined in Methods. These results are consistent with the quality of operations in our system. We find that we can create Bell states of two beryllium ions in the three-ion chain with fidelities of up to 97.8(4)%, as well as maximally entangled three-qubit Greenberger–Horne–Zeilinger states that also include the calcium ion with fidelities up to 93.8(5)%.

Prior to the next round of measurements, we re-initialize the Ca^+ qubit and the axial motional modes, which affect the performance of the multi-qubit gates. Fluorescence detection of Ca^+ scatters photons, which induces considerable heating of both the in-phase and ‘Egyptian’

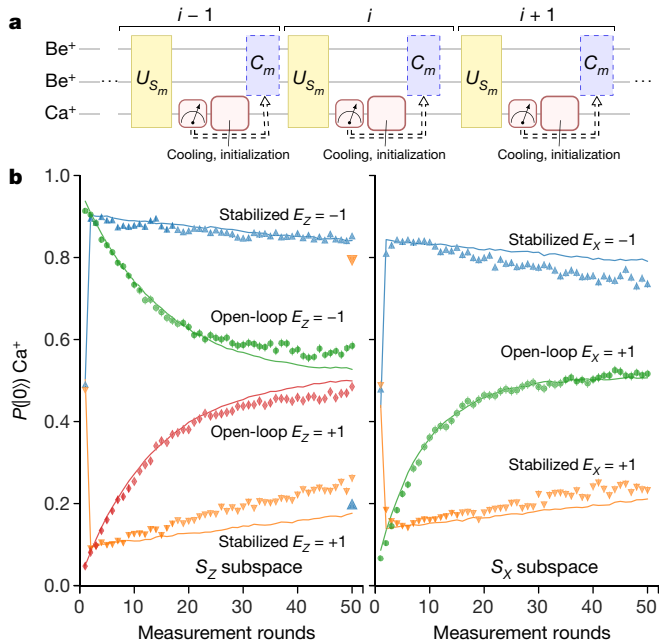


Fig. 2 | Repeated parity measurement and parity subspace stabilization. **a**, Illustration of the sequence used to stabilize a subspace, showing the use of cooling and feedback. Depending on the choice of stabilization, the unitary U_{S_m} is U_{S_z} or U_{S_x} , with C_m being the conditional feedback operation C_z or C_x , respectively. Recooling is performed using the Ca^+ ancilla. **b**, Ca^+ detection outcome probabilities as a function of repeated parity readouts for the S_z subspace. The circles (diamonds) show the outcomes as a function of repeated measurement rounds for the S_z subspace without feedback when initially preparing an $E_z = +1$ ($E_z = -1$) Be^+ input state. The upward- and downward-pointing triangles show the results obtained with the feedback conditioned on the Ca^+ results, stabilizing the two $E_z = +1$ and $E_z = -1$ subspaces. The larger upward-pointing (downward-pointing) triangles on the right show the Be^+ parity, as defined in the caption of Fig. 1, measured at the end of the $E_z = -1$ ($E_z = +1$) stabilization sequence. **c**, Similar data for stabilization of the S_x subspace. In this case, circles show the outcomes without feedback after preparing an $E_x = +1$ input state. Solid lines in **b** and **c** are produced using a simple Monte Carlo density matrix simulation using only one free parameter (see Methods for a discussion about the model and the discrepancies with measured data). Each data point is the result of 2,000 measurements. Error bars are derived from 1σ uncertainties originating from quantum projection noise and are smaller than markers.

modes of motion. Owing to the symmetry of our ion chain, the scattering does not heat the out-of-phase mode of motion of the two Be^+ ions (see Methods). However, the modes that are heated are sympathetically cooled by Ca^+ through EIT cooling followed by a few cycles of pulsed sideband cooling²⁷. We then optically pump the Ca^+ ion to $|0\rangle$, initializing it for the subsequent round of parity readout. Figure 2 shows the Ca^+ readouts for up to 50 sequential parity measurements performed on a pair of ions initially prepared in the $E_z = \pm 1$ subspace.

By introducing a feedback correction conditioned on the M_{S_z} or M_{S_x} outcomes using our flexible and low-latency classical control system (see Methods), we demonstrate the system's ability of stabilizing a parity subspace. The correction operation for S_z is $C_z = -I_1 \otimes X_2$, whereas for the S_x eigenspaces we implement $C_x = -I_1 \otimes Z_2$. In each case this requires a differential rotation of the Be^+ qubits. This is implemented by changing the potentials of the trap to shift the three-ion chain in the Be^+ Raman control laser beam so that one ion experiences twice the Rabi frequency of the other. At that position, we use a pulse that performs an X operation on one qubit while performing a $-I$ operation on the other. These pulses also produce undesired but stable a.c. Stark shifts, which we manage by updating the Raman laser differential phase for subsequent coherent operations. C_x consists of the C_z operation, with additional $R_{\pi/2}$ rotations around the Y axis before and after C_z .

Corrections are applied every time the Ca^+ measurement indicates the undesired subspace. Results from measurements with feedback stabilization are shown in Fig. 2b. Starting in one of the computational-basis states, we initially prepare an equal superposition of eigenstates with $E_z = \pm 1$ by applying a $\pi/2$ rotation to both beryllium ions. Linear fits to the closed-loop data, apart from the first point, exhibit decay constants of around 0.3% per measurement round (exponential fits are not possible here owing to the lack of information about the long-time limit).

To obtain a figure of merit for the improvement achieved by this stabilization, we compare the decay rates γ of exponential fits to the open-loop data with that obtained from the linear fits. The comparison shows a clear improvement, of the order of $\gamma_{\text{open}}/\gamma_{\text{closed}} \approx 20$. The decay rates observed for the closed-loop data are of a level similar to what we can ascribe to a simulation that includes an independently measured level of leakage from the qubit subspace of the beryllium ions into neighbouring hyperfine levels due to spontaneous photon scattering (solid lines in Fig. 2b; see Methods). This leakage does not fully account for the decay—although in separate simulations, not shown here, we observe that introducing a gradual change of 0.06% per measurement round in the parity readout fidelity results in closer agreement. This change is within what might be produced by effects such as thermal cycling of the acousto-optic modulators that control the laser beam intensities. We note that leakage is expected to be similar to the rate of error due to Raman scattering within the qubit subspace. Although the gates in our experiments are not currently limited by leakage error, once other sources of error are eliminated^{23,28}, this error will become an important factor, which cannot be easily corrected by standard QEC techniques and may favour the use of ions without hyperfine structure^{5,29}.

We now proceed to stabilize the four Bell states $(|\Phi_{\pm}\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2})$ and $(|\Psi_{\pm}\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2})$ of two qubits using the two commuting operators S_z and S_x . To prepare these states, we start from an arbitrary input by sequentially measuring S_z and S_x and performing conditional feedback if the measured eigenvalues differ from the desired result. We apply the correction operations C_z and C_x after both stabilizers have been read out. Results of the stabilization of these states are shown in Fig. 3. Also shown are the Bell state fidelities, measured by sampling the state at fixed points in the sequence. These states are determined using a combination of coherent rotations and subsequent measurements performed directly on the Be^+ ions (see Methods). As a comparison with the stabilized states, we also show data for an unstabilized $|\Psi_{+}\rangle$ Bell state input, which was produced by a single two-qubit MS gate on the Be^+ ions followed by a $R_{\pi/2}(\phi_p)$ pulse on both qubits with the appropriate phase. We observe a mean fidelity of 0.731(4) after a single block of measuring both stabilizers and correction for all of the Bell states. This value drops to a mean of 0.613(4) after 25 cycles (because each cycle contains two stabilizer measurements, this amounts to 50 measurement rounds), thus exhibiting that entanglement is preserved in the system even after this extended sequence of operations. The highest fidelity achieved is for the singlet $|\Psi_{-}\rangle$ state, which may be due to its insensitivity to common-mode phase errors.

We have demonstrated the readout of two-qubit stabilizer operators on an ancilla qubit of a different species, which allows readout without crosstalk to the data qubits. Together with conditionally applied correction operations, this allowed the stabilization of parity subspaces and entangled quantum states over tens of measurement and feedback rounds. These experiments therefore show the general elements of stabilizer readout and correction required for QEC. Nevertheless, many improvements are needed for QEC to operate reliably. QEC requires the measurement of higher-weight stabilizer operators³⁰ and needs to be made compatible with the demands of fault tolerance³. In our system, the primary error source is the stabilizer readout operation; the coherence time of our beryllium qubits is longer than any sequence that we demonstrate. It is therefore critical to reduce errors in the basic operations. Our known error sources are primarily motional decoherence during the multi-species gates, decoherence of the calcium qubit during

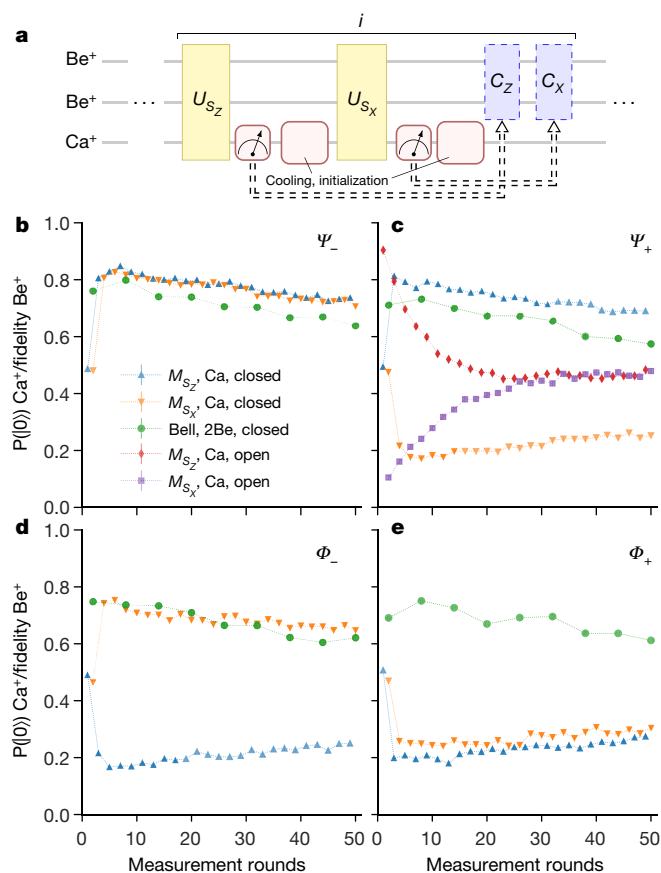


Fig. 3 | Bell state stabilization. **a**, Sequence for a measurement and feedback cycle used to generate and stabilize Bell states. M_{S_z} and M_{S_x} are performed sequentially, followed by sequential conditional feedback C_z and C_x . **b–e**, Evolution of the measured Ca⁺ outcome probabilities over 25 measurement and feedback cycles comprising 50 rounds of parity measurements. Also shown are the fidelities for the relevant states (green circles), measured after stopping the sequence at a fixed point and performing a set of measurements directly on beryllium. In **c** we also show the results from an open-loop experiment performed using the input state $|\Psi_+\rangle$ and repeatedly applying M_{S_z} without any feedback. The red diamonds show measurements of S_z and the purple squares correspond to S_x . Each data point is the result of 2,000 measurements. Error bars are derived from 1σ uncertainties originating from quantum projection noise and are smaller than markers.

the measurement block, and pulse calibration. Decoherence could be mitigated by applying gates at higher speeds using a higher laser intensity³¹, or by stabilizing laboratory parameters such as the magnetic field. Improved calibration could be achieved by reducing drifts and thermal effects, as well as introducing more frequent automated calibration routines. Indeed, we find that combining many different elements that include feedback operations applied probabilistically adds considerable complexity to the control and debugging of these systems, which motivates increased automation.

We anticipate extensions to this work in several directions. The control and measurement demonstrated here offer direct access to temporal correlations of multi-qubit systems (see Methods), with the potential to provide new insights into the evolution of quantum systems. The addition of conditional feedback opens up numerous opportunities for quantum state control. Although in this work we have focused on QEC, alternative applications include investigations of measurement-based quantum computing³², quantum gate teleportation³³ and quantum metrology⁴.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0668-z>.

Received: 17 May 2018; Accepted: 23 August 2018;
Published online 5 November 2018.

1. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
2. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493–R2496 (1995).
3. Gottesman, D. Theory of fault-tolerant quantum computation. *Phys. Rev. A* **57**, 127–137 (1998).
4. Leibfried, D. et al. Toward Heisenberg-limited spectroscopy with multiparticle entangled states. *Science* **304**, 1476–1478 (2004).
5. Terhal, B. M. Quantum error correction for quantum memories. *Rev. Mod. Phys.* **87**, 307–346 (2015).
6. Steane, A. M. Overhead and noise threshold of fault-tolerant quantum error correction. *Phys. Rev. A* **68**, 042322 (2003).
7. Cory, D. G. et al. Experimental quantum error correction. *Phys. Rev. Lett.* **81**, 2152–2155 (1998).
8. Chiaverini, J. et al. Realization of quantum error correction. *Nature* **432**, 602–605 (2004).
9. Reed, M. D. et al. Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482**, 382–385 (2012).
10. Barreiro, J. T. et al. An open-system quantum simulator with trapped ions. *Nature* **470**, 486–491 (2011).
11. Pfaff, W. et al. Demonstration of entanglement-by-measurement of solid-state qubits. *Nat. Phys.* **9**, 29–33 (2013).
12. Ristè, D. et al. Deterministic entanglement of superconducting qubits by parity measurement and feedback. *Nature* **502**, 350–354 (2013).
13. Cramer, J. et al. Repeated quantum error correction on a continuously encoded qubit by real-time feedback. *Nat. Commun.* **7**, 11526 (2016).
14. Sun, L. et al. Tracking photon jumps with repeated quantum non-demolition parity measurements. *Nature* **511**, 444–448 (2014).
15. Monz, T. et al. Realization of a scalable Shor algorithm. *Science* **351**, 1068–1070 (2016).
16. Schmidt, P. O. et al. Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
17. Tan, T. R. et al. Multi-element logic gates for trapped-ion qubits. *Nature* **528**, 380–383 (2015).
18. Barrett, M. D. et al. Sympathetic cooling of ⁹Be⁺ and ²⁴Mg⁺ for quantum logic. *Phys. Rev. A* **68**, 042302 (2003).
19. Home, J. P. et al. Complete methods set for scalable ion trap quantum information processing. *Science* **325**, 1227–1230 (2009).
20. Hume, D. B., Rosenband, T. & Wineland, D. J. High-fidelity adaptive qubit detection through repetitive quantum nondemolition measurements. *Phys. Rev. Lett.* **99**, 120502 (2007).
21. Langer, C. et al. Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).
22. Ballance, C. J. et al. Hybrid quantum logic and a test of Bell's inequality using two different atomic isotopes. *Nature* **528**, 384–386 (2015).
23. Gaebler, J. P. et al. High-fidelity universal gate set for ⁹Be⁺ ion qubits. *Phys. Rev. Lett.* **117**, 060505 (2016).
24. Roos, C. F. et al. Experimental demonstration of ground state laser cooling with electromagnetically induced transparency. *Phys. Rev. Lett.* **85**, 5547–5550 (2000).
25. Sørensen, A. & Mølmer, K. Entanglement and quantum computation with ions in thermal motion. *Phys. Rev. A* **62**, 022311 (2000).
26. Lee, P. J. et al. Phase control of trapped ion quantum gates. *J. Opt. B* **7**, S371–S383 (2005).
27. Monroe, C. et al. Resolved-sideband raman cooling of a bound atom to the 3D zero-point energy. *Phys. Rev. Lett.* **75**, 4011–4014 (1995).
28. Ballance, C. J., Harty, T. P., Linke, N. M., Sepiol, M. A. & Lucas, D. M. High-fidelity quantum logic gates using trapped-ion hyperfine qubits. *Phys. Rev. Lett.* **117**, 060504 (2016).
29. Brown, N. C. & Brown, K. R. Comparing Zeeman qubits to hyperfine qubits in the context of the surface code: ¹⁷⁴Yb⁺ and ¹⁷¹Yb⁺. *Phys. Rev. A* **97**, 052301 (2018).
30. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
31. Schäfer, V. et al. Fast quantum logic gates with trapped-ion qubits. *Nature* **555**, 75–78 (2018).
32. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
33. Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).

Acknowledgements We thank D. Kienzler, L. de Clercq and D. Nadlinger for contributions to the apparatus, J. Alonso for discussions and M. Grau and T.-L. Nguyen for reading the manuscript. We acknowledge support from the Swiss National Science Foundation through research grant 200020_165555 and through grant 51NF40-160591 from the National Centre of Competence in Research for Quantum Science and Technology (QSIT). K.K.M. is supported by an ETH Zürich Postdoctoral Fellowship. This research is partly based on work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the US Army Research Office grant W911NF-16-1-0070. The views and conclusions contained herein are those of the authors and should not be interpreted as

necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA or the US Government. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the US Army Research Office.

Reviewer information *Nature* thanks B. Blinov and T. Northup for their contribution to the peer review of this work.

Author contributions Experimental data were taken by V.N., M.M. and K.K.M., using an apparatus built primarily by M.M., V.N., H.-Y.L. and C.F. V.N. and K.K.M. performed the data analysis, and M.M., K.K.M. and J.P.H. performed the

modelling. V.N., M.M., K.K.M. and J.P.H. wrote the manuscript, with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0668-z>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to V.N. or M.M. or J.P.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Magnetic-field-insensitive beryllium qubit. Beryllium qubits are stored in the $|0\rangle \equiv |F=1, m_F=+1\rangle$ and $|1\rangle \equiv |F=2, m_F=0\rangle$ states of the hyperfine ground-state structure, where F and m_F are the total angular momentum and its projection along the quantization axis, respectively. The transition frequency is insensitive to first-order fluctuations around the applied magnetic field of 119.45 G. As a result, we obtain coherence times higher than 1 s for a single qubit. The internal-state preparation of the beryllium ions is performed by optical pumping to the $|F=2, m_F=2\rangle$ state, followed by a coherent Raman transition that transfers population to $|0\rangle$. Following the experimental sequence described in the main text, direct readout of beryllium is performed by first using two Raman transfer pulses, $|0\rangle \rightarrow |F=2, m_F=2\rangle$ and $|1\rangle \rightarrow |F=1, m_F=-1\rangle$, and subsequently illuminating the ions with resonant light on the $|S_{1/2}, F=2, m_F=2\rangle \rightarrow |P_{3/2}, F=3, m_F=3\rangle$ transition³⁴. Photons scattered by the ions are detected by a photomultiplier tube, resulting in distributions with a mean of 25 (0.2) counts per ion for the bright (dark) state. Owing to impure polarization causing optical pumping, a departure from Poissonian distributions is seen, resulting in an error of around 0.5% per ion in state detection. The beryllium ions are always read out simultaneously. We determine the probabilities of $P(|00\rangle)$, $P(|11\rangle)$ and $P(|01\rangle) + P(|10\rangle)$ by fitting a sum of three Poissonian distributions to the full histogram of detection counts.

Ion crystal and normal modes. Experiments are performed in a segmented linear Paul trap with a minimum ion–electrode distance of 185 μm . The ion chain used in the experiments is a three-ion chain ($\text{Be}^+ - \text{Ca}^+ - \text{Be}^+$) with calcium in the centre, which is formed and maintained using standard re-ordering techniques after every 50 experimental sequences³⁵. This allows the ions to recover from occasional re-ordering events due to background gas collisions. The ions exhibit three axial normal modes, which are schematically illustrated in Extended Data Fig. 1. For a harmonic potential, the calcium ion is decoupled from the mode in which the beryllium ions oscillate out of phase, which makes this mode unsuitable for multi-species gate operations. We choose the lowest-frequency motional mode (with a frequency of 1.56 MHz), in which all ions oscillate in-phase for the MS gates used in the parity measurement. For this mode, the Lamb–Dicke parameter is 0.13 for the beryllium ions and 0.07 for the calcium ion. Before each experimental run we cool the crystal using first 3 ms of far-detuned and then 600 μs of near-resonant Doppler cooling on Be^+ and Ca^+ simultaneously. This is followed by Ca^+ EIT cooling optimized on the lowest radial mode at 2.5 MHz, Be^+ Raman sideband cooling on the two highest-frequency axial motional modes, and finally Ca^+ sideband cooling on the lowest-frequency mode.

Once every several minutes we observe that the ion crystal enters a highly excited motional state, probably owing to background gas collisions with the light beryllium ions; we refer to this process as ‘decrySTALLIZATION’. This results in loss of fluorescence and possible loss of an ion without rapid intervention, which is mitigated by monitoring the fluorescence emitted by the Be^+ ions during Doppler cooling for each experimental shot. If it is below a threshold for more than a few shots, then the trap axial frequency is immediately reduced to 400 kHz, resulting in stronger radial confinement, and a loop of far-detuned and near-resonant Doppler cooling is repeated until the fluorescence returns to normal or a timeout is reached. In the former case the experiment is seamlessly re-run without user intervention. The loop takes 50–200 ms to complete. With automatic re-crystallization a three-ion crystal can remain trapped for 4–6 h at a time, with beryllium hydride formation being the dominant cause of ion loss.

Qubit control: beryllium and calcium. Be^+ qubit operations are mediated by pairs of Raman beams that are near-resonant with the qubit transition frequency of 1.2 GHz. For single-qubit operations, we use a pair of beams that are co-propagating as they enter the trap (the ‘co’ beam pair in Extended Data Fig. 1), such that any path-length fluctuations are common to both. This ensures that the difference phase, which sets the phase in the qubit frame of reference, exhibits minimal drift. The wavevector difference between these two beams is nearly zero, which means that they cannot mediate spin-motion coupling. For this reason, a second beam path involving a pair of Raman beams entering the apparatus at 90° to one another, with their difference wavevector aligned along the trap axis, is used for the MS entangling gates. Owing to the different optical paths of the beams, air currents and slight thermal drifts result in submicrometre-scale path-length fluctuations that are not common to both beams, and hence the phase ϕ_b discussed in the main text can drift. This is a known problem with such free-space beam geometries²⁶. We take advantage of the fact that these drifts are slow on the 100- μs timescales associated with our individual unitaries to enclose the gates using the ‘90’ beam pair (the MS gates and Be^+ π pulse; Extended Data Fig. 1) in a Ramsey sequence using the same pair, which serves to cancel this phase sensitivity.

The calcium qubit states are $|0\rangle \equiv |^2S_{1/2}, m_J=+1/2\rangle$ and $|1\rangle \equiv |^2D_{5/2}, m_J=+3/2\rangle$, for which the transition frequency is first-order-dependent on the magnetic field, resulting in a much shorter coherence time of 1.5 ms. The internal-state preparation of calcium ions is performed by frequency-selective optical pumping³⁶. The Ca^+ qubit is controlled using a single 729-nm laser propagating at 45° with respect

to the trap axis. State readout is performed by illuminating the ion with a resonant beam on the $|^2S_{1/2}\rangle \rightarrow |^2P_{1/2}\rangle$ transition for 200 μs . Photons scattered by the ion are detected on a photomultiplier tube, resulting in distributions with a mean of 21 (2) counts per ion for the bright (dark) state. State discrimination is performed by comparing the number of detected photons with a threshold that is independently obtained by fitting Poisson distributions to the reference data.

Phase updates and reset. The laser pulse frequencies, amplitudes and phases are controlled by radio-frequency synthesis³⁷. Each individual frequency used in the experiment is initialized to zero phase at a certain point before the coherent operations performed on the ions. The phase of each pulse is referenced to a ‘reference’ time t_{start} . Each radio-frequency tone (with angular frequency ω_{rf}) applied to the acousto-optic modulators has a phase that evolves as $\phi = \omega_{\text{rf}}(t - t_{\text{start}})$. Neglecting a.c. Stark shifts during the gates, the phases of the carrier pulses rotate with the qubit frame, and any additional phase added to particular pulses corresponds to the relevant phase in the qubit frame (that is, the phases written on carrier pulses in Fig. 1a). Although this scheme ensures that all carrier pulses from a particular beam or beam pair maintain fixed relative phase relationships, it generally does not automatically preserve the relative phase relationship between radio-frequency pulses detuned from the carrier. In our MS gates, for example, a common offset frequency introduced to both sidebands (to compensate for Stark shifts occurring during the gate) results in a relative phase that is dependent on the timing of this pulse relative to others in the sequence. As a result, simply repeating U_{S_Z} with the same pulse phases relative to t_{start} actually implements different unitaries at each cycle. To mitigate this, the reference time t_{start} is shifted to a fixed time before every U_{S_Z} stabilizer readout pulse block on each occasion that the latter is applied. It is then reset back to the experiment start time before any further carrier rotations (using the ‘co’ pair) on the qubits. Because the S_Z stabilizer readout block is diagonal in the computational basis, the phases of the pulses in this block do not need to be referenced to the rest of the sequence.

In addition to the time-reference shifting, the Stark shifts due to the U_{S_Z}, U_{S_X}, C_X and C_Z operations in the feedback experiments are dynamically compensated by adjusting future gate phases according to the previously conducted operations. Unlike the phase accumulation described earlier, which is hard-coded, these calculations are carried out rapidly in our embedded software, and the results, such as the instructions for the feedback correction operations, are pushed to the radio-frequency hardware with 1–3 μs latencies. The logic for the feedback and phase calculations is implemented in software written in a high-level programming language (C++) running on a microprocessor that interfaces directly with a field-programmable gate array, allowing flexible reconfiguration of the system behaviour.

Fidelity estimation of single parity measurements. For a single round of parity state preparation and measurement, the measurement fidelity F_{ND} quoted in the main text is an estimate of how well the Ca^+ readout probabilities agree with the Be^+ input state parity probability distribution. The quantum state preparation fidelity F_{QSP} gives a measure of the mean conditional probability of the resulting Be^+ subspace corresponding to the measurement outcome obtained from the Ca^+ readout^{10,38}. Mathematically these are given by

$$F_{\text{ND}} = \left(\sqrt{p_{+1}^{\text{in}} p_{+1}^{\text{m}}} + \sqrt{p_{-1}^{\text{in}} p_{-1}^{\text{m}}} \right)^2$$

$$F_{\text{QSP}} = p_{+1\&1}^{\text{out}} + p_{-1\&0}^{\text{out}}$$

where $p_{+1(-1)}^{\text{in}}$ is the probability of finding the two Be^+ ions in the $E_Z = +1$ ($E_Z = -1$) subspace in a direct measurement performed on the input state (the violet stars in Fig. 1b show $p_{+1}^{\text{in}}, p_{-1}^{\text{in}}$). $p_{+1(-1)}^{\text{m}}$ ($p_{+1(-1)}^{\text{m}}$) is the probability of measuring Ca^+ in the dark (bright) state. $p_{+1\&1}^{\text{out}}$ ($p_{-1\&0}^{\text{out}}$) is the joint probability that after M_{S_Z} the beryllium state is in the $E_Z = +1$ ($E_Z = -1$) eigenspace and the calcium state is $|1\rangle$ ($|0\rangle$) (the blue circles and red squares in Fig. 1b show $p_{+1\&1}^{\text{out}}$ and $p_{-1\&0}^{\text{out}}$, respectively). The two joint probabilities are not independent, and the F_{QSP} error is calculated considering binomial statistics. F_{ND} and F_{QSP} are averages of F_{ND} and F_{QSP} , respectively, over all input state preparation phases.

Direct estimation of beryllium Bell state fidelity. To estimate beryllium state fidelities we measure correlations between the ions in three orthogonal bases. The first is the computational basis, which we obtain by a direct measurement of the two ions, by extracting the probability of finding the ions in the same state that we use to obtain $\langle S_Z \rangle$. For the $|\Phi_{\pm}\rangle$ states, we then apply $R_{\pi/2}$ pulses to both ions, with the phase of the pulse set such that two-ion populations observed in subsequent fluorescence detections allow the estimation of $\langle S_X \rangle$ and $\langle S_Y \rangle$ (this corresponds to estimating the contrast in parity oscillations from the two extrema). For the $|\Psi_{\pm}\rangle$ states, we precede the common $R_{\pi/2}$ pulse with a $C_Z = -I_1 \otimes X_2$ operation, which converts the states to one of the $|\Phi_{\pm}\rangle$ states, with the resulting analysis performed in the same manner as before. Although the use of the C_Z operation is not strictly necessary, it provides a simple set of diagnostics that we find easy to manage and debug. In the feedback experiments, it is important to ensure that the phase of

the pulses used in the analysis is updated according to the beryllium feedback operations that have been performed.

Correlations in Bell state stabilization. To better understand the Bell state stabilization data, we investigate the two-point correlations between successive Ca^+ readouts in the S_Z or S_X bases for the four Bell states, which provide useful diagnostic information on the feedback operations.

Each correlation datum is categorized on the basis of which correction operations have taken place between the two stabilizer measurements, and the subdivided data are shown in Extended Data Fig. 2. Ideally, for $S_Z(S_X)$, these correlations would have value 1 for a $C_X(C_Z)$ operation because this operation commutes with the stabilizer.

We observe that the correlations are on average about 10% lower with nominally commuting correction operations than in the case without feedback, which indicates that experimentally the $C_X(C_Z)$ operations do not commute with $U_{S_Z}(U_{S_X})$ as well as expected from calibration experiments, where errors of the level of 1% would be expected.

In a perfect implementation of measurement and feedback, aside from the first round of corrections, the C_Z correction operation is only applied when an additional source of error changes the subspace. In this case, the S_Z correlation between two detections with a C_Z operation between them should always be anti-correlated. However, when the dominant error is due to imperfect readout, the steady-state situation produces a 50% anti-correlation, because in half of all cases the correction operation is applied on a state that is uncorrupted. Because our data are primarily affected by errors introduced during the readout, for this correlation we see higher levels of 30%–50%.

Modelling of leakage and decoherence. We simulate the behaviour of our sequence of measurements using a simple model that allows us to quantify the main observed behaviours. This is implemented in a Monte Carlo simulation of the density matrix. The action of the measurement sequence itself is modelled by the ideal unitary operators $U_{S_Z|X}$, which map the parity information of the data qubits onto the ancilla. We mimic decoherence and Raman off-resonant scattering using two additional maps. The first is a depolarizing channel³⁰, which is applied as a map to the density operator after the application of the unitary, but before the measurement result is extracted. This is parameterized by a depolarization rate of γ_{dep} per measurement round. The measurement of the calcium ion is then performed using a projection. Following this, we add a leakage channel, which is implemented by a partial trace of one or both of the beryllium ions, applied probabilistically through the comparison of a random number with a reference drawn from the leakage probability distribution. This uses a leakage decay rate of γ_{leak} per ion per measurement round. We fix the leakage rate to $\gamma_{\text{leak}} = 0.3\%$, which was experimentally determined with independent measurements of the population remaining in the qubit subspace after applying the looped sequence with just one of the two Raman beams on at a time. The depolarization rate γ_{dep} is instead

adjusted for each dataset. In the open-loop data we adjust γ_{dep} to best reproduce the exponential decay of the state whereas for the closed-loop sets we use it to adjust the offset of the first few points, thus reflecting the infidelities of the feedback operations. We find for the open-loop S_Z measurement $\gamma_{\text{dep}} = 0.06$ ($\gamma_{\text{dep}} = 0.07$) for $E_Z = -1$ ($E_Z = +1$) and $\gamma_{\text{dep}} = 0.11$ for $E_X = +1$. We attribute the higher value for the $E_X = +1$ measurements to the higher susceptibility to phase fluctuations. For the closed-loop sets we find $\gamma_{\text{dep}} = 0.10$ ($\gamma_{\text{dep}} = 0.10$) for $E_Z = -1$ ($E_Z = +1$) and $\gamma_{\text{dep}} = 0.15$ ($\gamma_{\text{dep}} = 0.16$) for $E_X = +1$ ($E_X = -1$). These higher values are associated with imperfections in the correction pulses, as described in the previous section. We note that the errors that occur in our experimental system are probably not well modelled by depolarizing noise. Nevertheless in this complex system it gives a simple single-parameter adjustment that accounts for decoherence.

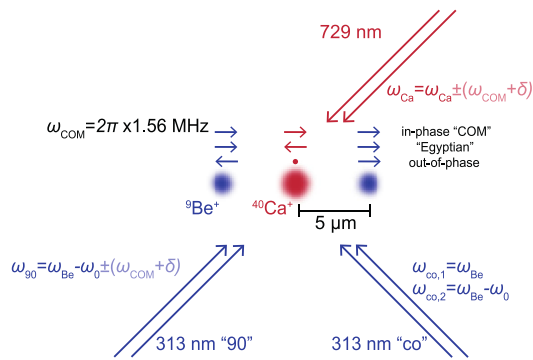
As can be seen in Fig. 2, the simulation results that include only leakage and depolarization do not fully reproduce the observed dynamics. By adding a gradual drift in the readout fidelity of the parity measurement, we are able to simultaneously produce a good match to the data for all datasets. Such a drift could arise experimentally from a number of sources. We observe gradual changes in pulse amplitudes over the sequence length, as well as a change in the photomultiplier-tube counts. Combined, these problems are consistent with a parity detection bias that increases over the duration of the sequence to 3%–5%.

Using this simulation we can also examine the behaviour of correlations between feedback events, as described in the previous section. In Extended Data Fig. 3 we plot the simulated and theoretical probabilities of two feedback events being applied in subsequent shots. We see that the simulation reproduces the experimental results reasonably well, which verifies the arguments made in the previous section regarding correlations.

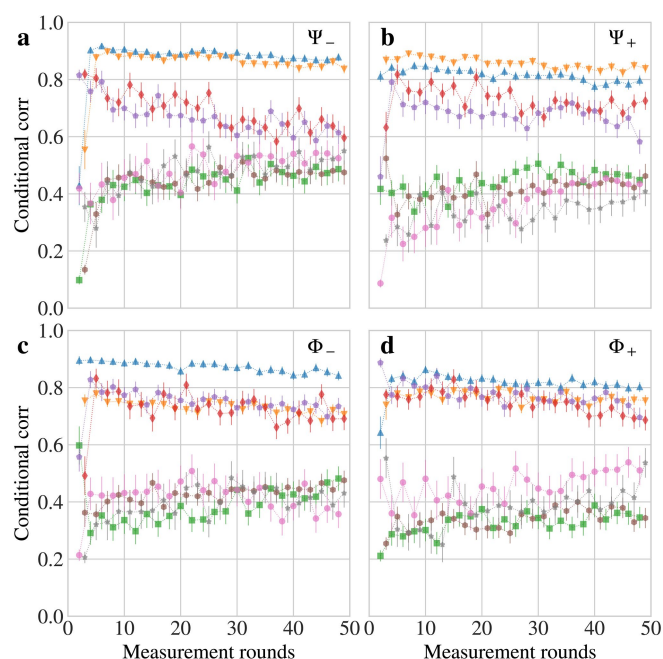
Data availability

The data generated and analysed during this study are available from the corresponding authors upon reasonable request.

34. de Clercq, L. E. et al. Parallel transport quantum logic gates with trapped ions. *Phys. Rev. Lett.* **116**, 080502 (2016).
35. Home, J. P. Chapter 4 – quantum science and metrology with mixed-species ion chains. *Adv. Atom. Mol. Opt. Phys.* **62**, 231–277 (2013).
36. Kienzler, D. et al. Quantum harmonic oscillator state synthesis by reservoir engineering. *Science* **347**, 53–56 (2015).
37. Keith, B., Negnevitsky, V. & Zhang, W. Programmable and scalable radio-frequency pulse sequence generator for multi-qubit quantum information experiments. Preprint at <http://arxiv.org/abs/1710.04282> (2017).
38. Ralph, T. C., Bartlett, S. D., O'Brien, J. L., Pryde, G. J. & Wiseman, H. M. Quantum nondemolition measurements for quantum information. *Phys. Rev. A* **73**, 012113 (2006).

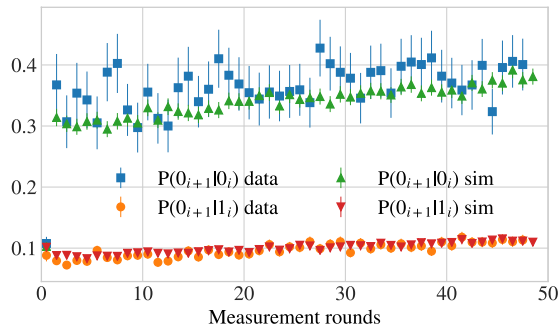


Extended Data Fig. 1 | Ion crystal and beam configuration. Three co-trapped ions in a single harmonic potential well are addressed by Raman beam pairs for Be^+ and a 729-nm beam for Ca^+ . The laser frequencies (written in solid text) correspond to those applied for single-qubit operations, resonant with the qubit transition (here ω_{Ca} is the optical qubit frequency, ω_{Be} is an optical frequency near 313 nm, ω_0 is the beryllium microwave qubit frequency, and $\omega_{\text{co},1}$ and $\omega_{\text{co},2}$ are the laser frequencies of the co-propagating Raman beams). The faint text represents modified frequencies used to address the sidebands that are associated with the in-phase centre-of-mass (COM) mode of the three-ion crystal at 1.56 MHz, at a small detuning δ to drive the entangling MS gates. Arrows above the ions indicate the normal-mode displacement directions for this mode, as well as the 'Egyptian' (4.20 MHz) and the out-of-phase (4.11 MHz) axial modes.



Extended Data Fig. 2 | Experimental Bell measurement correlations.

Correlations between pairs of successive M_{S_Z} measurement operations, categorized by the feedback that took place between them. Values of 1 (0) represent perfect correlation (anti-correlation). Upward-pointing (downward-pointing) triangles show the S_Z (S_X) correlations where no feedback occurred; thus, for S_Z the sequence is $S_Z-S_X-S_Z$. Pentagons (diamonds) are data obtained with the application of C_X (C_Z), which nominally commutes with S_Z (S_X); for example, the sequence $S_Z-S_X-C_X-S_Z$. Squares (hexagons) are correlations between two measurements in which C_Z (C_X)—namely, a correction that anti-commutes with the operator of interest—took place. Circles and stars indicate data for which both C_X and C_Z occurred. The plots use the same raw data as Fig. 3, and uncertainties reflect projection noise; the uncertainties fluctuate noticeably owing to the varying numbers of events considered for each point, which corresponds to a subset of the full dataset.



Extended Data Fig. 3 | Simulated subspace measurement correlations.

By defining $P(0_i)$ as the probability of applying a feedback operation in the i th measurement round, we plot the conditional probability of feeding back on the system twice in a row ($P(0_{i+1}|0_i)$) or just once ($P(0_{i+1}|1_i)$) for both the experimental data and the simulation ('sim'). Uncertainties for $P(0_{i+1}|0_i)$ are much larger owing to the rarity of these events (an average of 100 events over 10,000 simulated points).

Flight of an aeroplane with solid-state propulsion

Haofeng Xu¹, Yiou He², Kieran L. Strobel¹, Christopher K. Gilmore¹, Sean P. Kelley¹, Cooper C. Hennick¹, Thomas Sebastian³, Mark R. Woolston³, David J. Perreault² & Steven R. H. Barrett^{1*}

Since the first aeroplane flight more than 100 years ago, aeroplanes have been propelled using moving surfaces such as propellers and turbines. Most have been powered by fossil-fuel combustion. Electroaerodynamics, in which electrical forces accelerate ions in a fluid^{1,2}, has been proposed as an alternative method of propelling aeroplanes—without moving parts, nearly silently and without combustion emissions^{3–6}. However, no aeroplane with such a solid-state propulsion system has yet flown. Here we demonstrate that a solid-state propulsion system can sustain powered flight, by designing and flying an electroaerodynamically propelled heavier-than-air aeroplane. We flew a fixed-wing aeroplane with a five-metre wingspan ten times and showed that it achieved steady-level flight. All batteries and power systems, including a specifically developed ultralight high-voltage (40-kilovolt) power converter, were carried on-board. We show that conventionally accepted limitations in thrust-to-power ratio and thrust density^{4,6,7}, which were previously thought to make electroaerodynamics unfeasible as a method of aeroplane propulsion, are surmountable. We provide a proof of concept for electroaerodynamic aeroplane propulsion, opening up possibilities for aircraft and aerodynamic devices that are quieter, mechanically simpler and do not emit combustion emissions.

Electroaerodynamics (EAD) is a means of generating propulsive forces in fluids^{1,2}. Ions generated in the ambient fluid and under the influence of an applied electric field are accelerated by the Coulomb force. These ions collide with neutral molecules and couple the momentum of the accelerated ions with that of the bulk fluid; the result is an ionic wind that produces a thrust force in the opposite direction to ion flow. In our device, we generate ions using a corona discharge. A corona discharge is a self-sustaining atmospheric discharge that is induced by the application of a constant high electric potential across two asymmetric electrodes; high electric fields near the smaller electrode accelerate electrons and produce a cascade of ionization by successive electron collisions with neutral molecules⁸.

Electroaerodynamic propulsion is a method of manipulating and moving fluids without any need for moving surfaces, making it attractive for a number of applications. For example, the concepts of electrohydrodynamics (where the neutral fluid is water) and electroaerodynamics (where the neutral fluid is air) have been investigated for heat-transfer enhancement^{9,10} and ion drag pumps².

The additional advantages of being nearly silent and producing no combustion emissions make EAD particularly attractive as a propulsion system for aeroplanes. It could potentially mitigate the harmful impact of current aeroplane propulsion systems on the environment, particularly given the anticipated growth of urban drone usage and its associated noise impacts; EAD could enable the design of quieter, smaller aircraft that interact more closely and innocuously with the urban environment. Its solid-state nature could also enable miniaturization to an extent not possible with conventional propulsion¹¹. However, the feasibility of EAD as a method of propulsion is confronted by the challenges of producing sufficient thrust, while achieving low aircraft drag and weight.

Although there have been a number of design proposals for heavier-than-air EAD^{6,7} aeroplanes, no such aircraft has yet flown. No other kind of aeroplane with a solid-state propulsion system has flown

either (unless supersonic ramjets were to be considered in this category, although the intakes have moving surfaces). This Letter describes the powered flight of such an aeroplane.

A viable propulsion system must produce sufficient thrust without a large weight or drag penalty. This sets limits both on the power requirements (that is, the thrust-to-power ratio) and on the frontal area (that is, thrust density) of the EAD system.

The first of these potential limitations, the thrust-to-power ratio, can be estimated using a one-dimensional steady-state parallel-plate model first presented by Chattock¹ and later experimentally confirmed by Stuetzer², and Christensen and Møller³. It consists of two electrodes, across which there is a sufficiently high electric potential, resulting in the formation of a small ionization region near the anode and a dominant unipolar conduction region of positively charged ion drift from anode to cathode. The electrostatic behaviour is governed by Gauss's law:

$$\frac{dE}{dx} = -\frac{d^2V}{dx^2} = \frac{\rho}{\epsilon_0}$$

where E is the electric field strength, V is the electric potential, ρ is the charge density, ϵ_0 is the permittivity of free space, and x is the distance along the coordinate direction. Under an inviscid fluid assumption, the fluid momentum equation is governed by the Euler equation: $dp/dx = \rho E$, where p is the pressure. The current density $j = \rho(\mu E + v_0)$, where μ is the ion mobility, is composed of the EAD current due to the applied electric field and the convective drift due to the freestream velocity v_0 .

Substituting and integrating the above equations gives an analytical equation for the thrust-to-power ratio, which is a measure of static thrust efficiency and an important figure of merit for propulsion systems:

$$\frac{T}{P} = \frac{(\rho \bar{E})LA}{V(jA)} = \frac{\rho \bar{E}L}{V\rho(\mu \bar{E} + v_0)} = \frac{1}{(\mu \bar{E} + v_0)}$$

where T is thrust, P is power, \bar{E} is the average electric field strength, A is the thruster area, and L is the inter-electrode distance. A low average electric field strength \bar{E} gives a higher thrust-to-power ratio, but too low an electric field strength results in no corona inception. Thus there is a limit to the achievable thrust-to-power ratio using the corona discharge. Masuyama and Barrett⁵ found experimentally that a thrust-to-power ratio as high as 50 N kW^{-1} was achievable at the laboratory scale, comparable to that of conventional propulsion, where typical performance is about 3 N kW^{-1} for jet engines¹² and about 50 N kW^{-1} for helicopter rotors¹³.

A parametric study by Gilmore and Barrett⁷ quantified the thrust density of EAD propulsors under a number of different parallel and staged electrode configurations. A trade-off between thrust-to-power ratio and thrust density was identified. At a fixed operating voltage and geometry, an increase in the thrust-to-power ratio results in a decrease in thrust density, and vice versa. The experiments suggested that a thrust density of 3 N m^{-2} and a thrust-to-power ratio of 6.25 N kW^{-1} could be simultaneously achieved with a two-staged configuration of four sets of parallel electrodes. The same configuration is used here (Extended Data Fig. 1).

This level of performance suggested that steady-level flight of a fixed wing unmanned aircraft might be feasible, but at the limit of what is

¹Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA. ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ³Massachusetts Institute of Technology Lincoln Laboratory, Cambridge, MA, USA. *e-mail: sbarrett@mit.edu

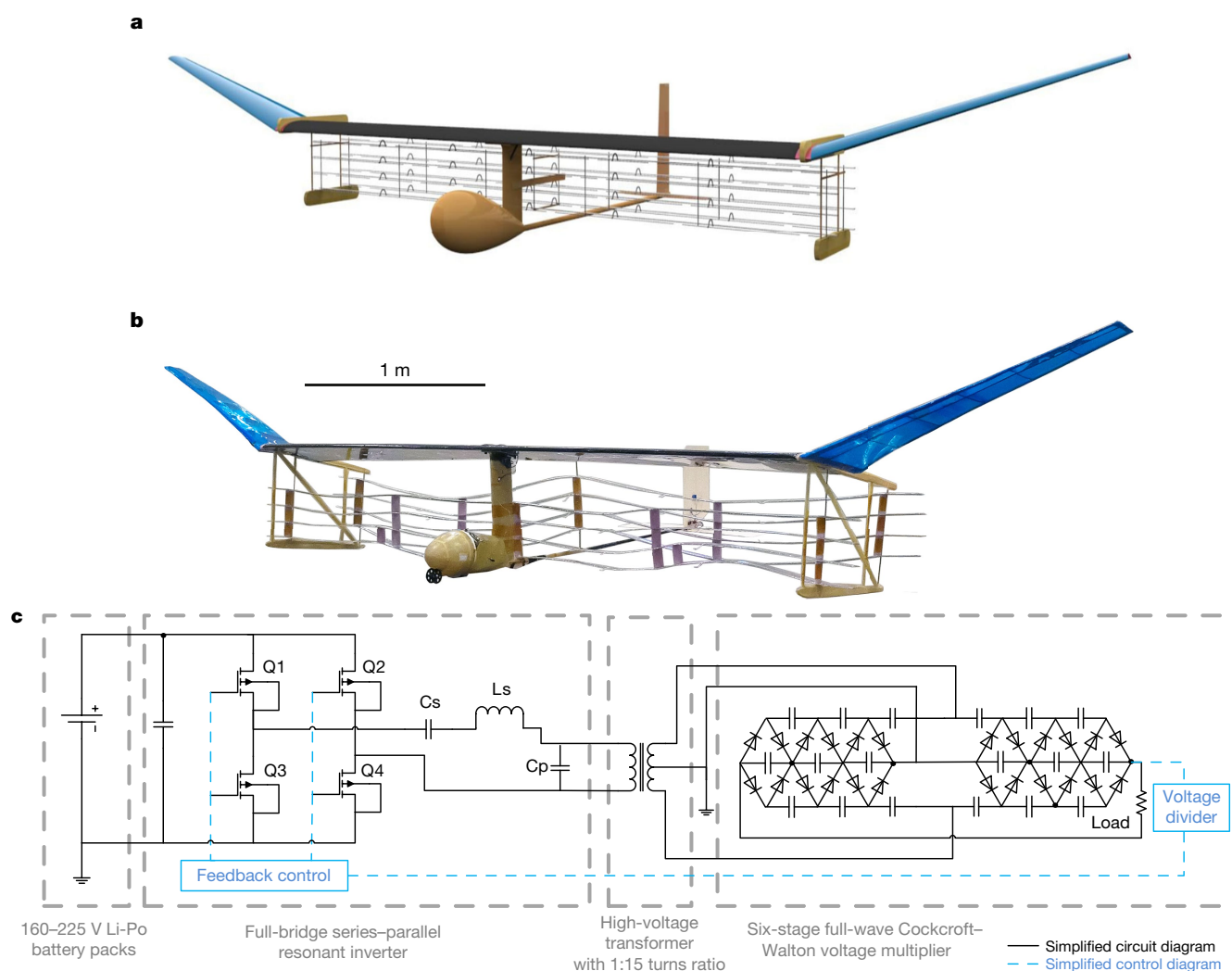


Fig. 1 | Aeroplane design. **a**, Computer-generated rendering of the EAD aeroplane. **b**, Photograph of actual EAD aeroplane (after multiple flight trials). **c**, Architecture of the high-voltage power converter (HVPC). The HVPC consists of three stages: a series–parallel resonant inverter that converts 160–225 V direct current to a high-frequency alternating current, a high-voltage transformer that steps up the alternating-current

voltage, and a full-wave Cockcroft–Walton multiplier that rectifies the high-frequency alternating current back to direct current. The resonant converter uses transformer parasitics (including transformer capacitance) as part of the resonant tank. The three stages contribute a voltage gain of about $2.5\times$, $15\times$ and $5.6\times$.

technologically possible using current materials and power electronics technology. It was therefore necessary to systematically search the possible design space for a feasible aircraft design—if one were to exist. We used a geometric programming optimization method to find the most viable size and power for the prototype aircraft design, finding

that steady-level flight was just achievable. Geometric programming has been applied to design optimization of conventional aircraft systems¹⁴. It is able to efficiently find the global optimum of a set of design variables for a certain objective function, given a series of inequality constraints. We chose aircraft wingspan as the objective function to be

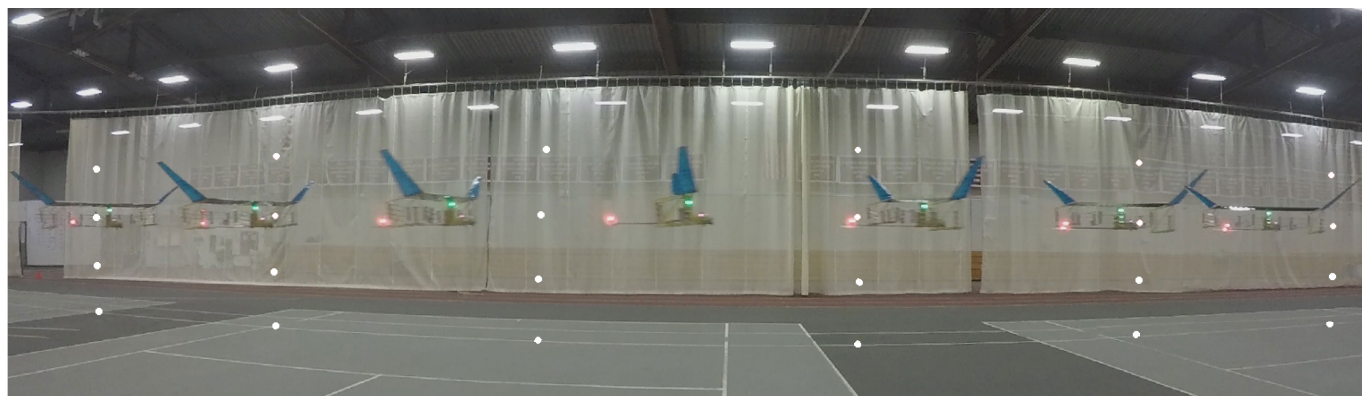


Fig. 2 | Time-lapse image of the EAD aeroplane in flight. White reference markers (spots) are spaced 5 m apart horizontally and 1 m apart vertically. All subsequent results are presented in a Cartesian coordinate

system with the x axis in the flight direction, z axis upwards, and y axis pointing away from the camera.

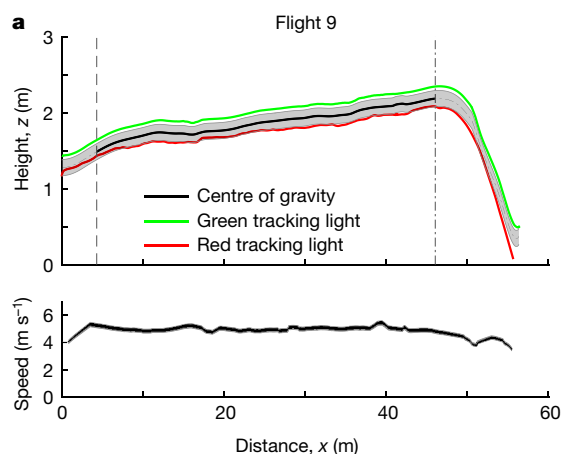


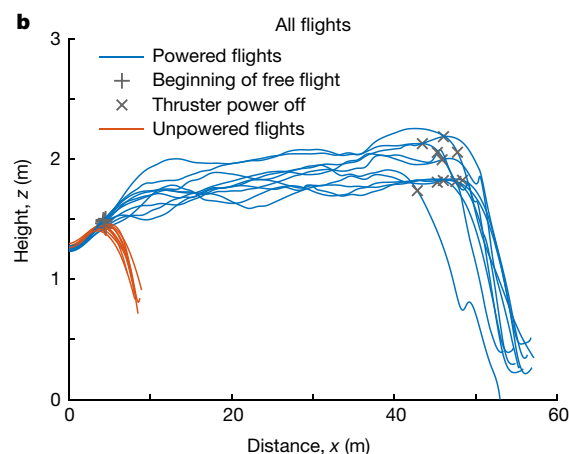
Fig. 3 | Flight trajectories. **a**, Flight trajectory (top) and speed profile (bottom) for a single flight (number 9): the position of the centre of gravity between launch (vertical dashed line) and propulsion system power-off (vertical dash-dotted line) is shown by the solid black line.

minimized, which corresponds with low weight, low electrical power supply requirements, and low development time, risk and cost. The geometric programming design optimization model incorporated aerodynamic, structural, EAD and power electronics models.

The design optimization was able to find a feasible solution at a wing-span of 5 m with a design weight of 2.45 kg, flight velocity of 4.8 m s^{-1} , thrust of 3.2 N, and electrical power requirement of 600 W (Fig. 1, Extended Data Table 1). We used a conventional tail with elevator and rudder for stability and trim. However, future EAD aircraft without moveable control surfaces are possible.

To produce a sustained thrust of 3 N at the 40 kV design point and estimated thrust-to-power ratio of 6.25 N kW^{-1} , a power system capable of delivering 500 W of output power was required, but at a very low weight. These weight constraints necessitated the design and construction of both a custom battery stack and a custom high-voltage power converter (HVPC) which stepped up the battery voltage to 40 kV. Our HVPC achieved a specific power of 1.2 kW kg^{-1} , 5–10 times higher than conventional power supplies at this voltage and power¹⁵. This was enabled by careful converter topology selection (Fig. 1c), by design optimization and by operating at a high switching frequency between 500 kHz and 700 kHz to minimize the weight of the capacitors and the magnetics^{16,17}.

We performed ten flights with the full-scale experimental aircraft at the MIT Johnson Indoor Track (Fig. 2). Owing to the limited length of the indoor space (60 m), we used a bungeed launch system to accelerate the aircraft from stationary to a steady flight velocity of 5 m s^{-1} within 5 m, and performed free flight in the remaining 55 m of flight space. We also



The estimated position tracking error is shown in grey (see Methods). **b**, Trajectories for all ten powered flights (blue) and ten unpowered glides (orange). The steady segment of flight covered a distance of 40–45 m with a duration of 8–9 s.

performed ten unpowered glides with the thrusters turned off, in which the aeroplane flew for less than 10 m. We used cameras and a computer vision algorithm to track the aircraft position and determine the flight trajectory.

All flights gained height over the 8–9 s segment of steady flight, which covered a distance of 40–45 m (Fig. 3). The average physical height gain of all flights was 0.47 m (Fig. 4a). However, for some of the flights, the aircraft velocity decreased during the flight. An adjustment for this loss of kinetic energy (Fig. 4b) results in an energy equivalent height gain, which is the height gain that would have been achieved had the velocity remained constant. This was positive for seven of the ten flights, showing that better than steady-level flight had been achieved in those cases.

We have thus shown that the feasibility of EAD as a method of propulsion is not prohibited by previously identified limitations in thrust-to-power ratio and thrust density. In this proof of concept for this method of propulsion, the realized thrust-to-power ratio was 5 N kW^{-1} , which is of the order of conventional aeroplane propulsion methods such as the jet engine. However, the overall efficiency was lower than typically achieved by conventional propulsion (and was not the objective here given the limited indoor test area for an uncertified aeroplane). Using the in-flight HVPC power-draw measurements, and confirmation using dynamic wind tunnel experiments, we estimate that the thruster produced 3.2 N of thrust, giving an estimate of overall efficiency:

$$\eta_{\text{overall}} = \frac{T|\vec{v}|}{P_{\text{input}}} = \frac{3.2 \text{ N} \times 4.8 \text{ m s}^{-1}}{600 \text{ W}} = 2.56\%$$

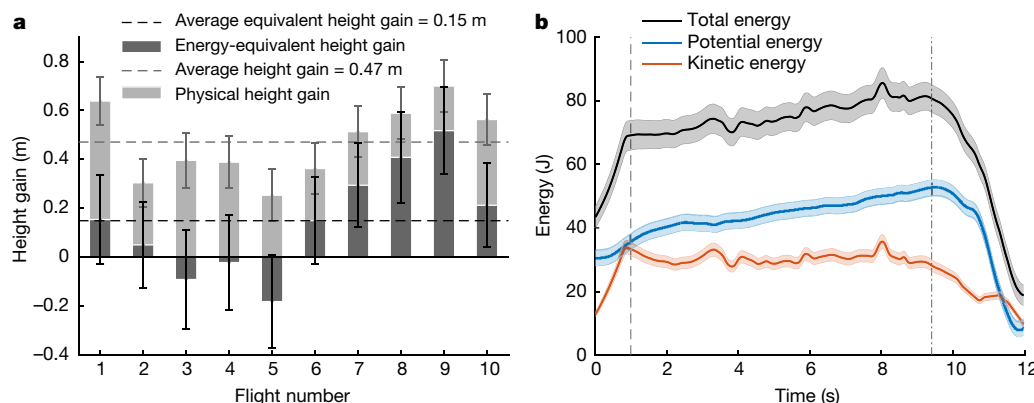


Fig. 4 | Steady-level flight. **a**, The physical height gain was positive for all flights and the energy-equivalent height gain, which adjusts for the loss of kinetic energy during the flight, was positive for seven flights. Zero energy-equivalent height gain indicates steady-level flight. **b**, The variation in kinetic energy, potential energy and total energy (the sum of kinetic and

potential energies) during a particular flight: the potential energy increases substantially, while the kinetic energy remains approximately constant or decreases slightly. Vertical dashed and dash-dotted lines as in Fig. 3a. See Methods for details of error estimation.

which is defined as the thrust power delivered to the aircraft divided by the input electrical power P_{input} ; T is the magnitude of the thrust force and \bar{v} is the average velocity of the aircraft in the direction of thrust. To achieve longer range and endurance for practical applications, future research should aim for increasing overall efficiency.

This prototype aircraft was designed by optimizing for small wing-span and low power requirement, which resulted in low overall efficiency, largely due to a low flight speed. This explains the discrepancy between the apparently comparable thrust-to-power performance and the poor overall efficiency performance (which is also a function of the airframe and flight conditions). Using the same EAD performance models and the same aircraft design models, but optimizing for maximum efficiency instead of minimum wingspan, suggests that aircraft with 5% overall efficiency are readily achievable with current technology. This aircraft would be larger, fly faster and require a more powerful HVPC. This design may be useful for applications where low noise and no moving parts are critical, but it is not yet competitive against conventional aeroplanes at similar scale in metrics such as range, endurance and payload capacity.

Further technology improvements in EAD propulsion are needed to increase overall efficiency. Areas include exploring alternative ionization regimes (which could increase practically achievable thrust density and reduce viscous drag from EAD electrodes), designing ways to increase HVPC specific power and efficiency, and integrating the propulsion system with the airframe to reduce the overall aerodynamic power requirements for flight. In the limiting case, one-dimensional EAD theory suggests that overall efficiency for an idealized thruster could be as high as 50%. This would require that a high thrust-to-power ratio be coupled with a configuration that enables higher flight speeds without incurring overwhelming viscous drag^{3,5}.

Another remaining performance challenge is the achievable thrust density using EAD. Although we have shown that EAD thrust density is sufficient at the scale of unmanned aerial vehicles, where the available ratio of frontal area to weight is high, it is not currently sufficient for high-speed flight at the scale of commercial aviation: the area thrust density of our aeroplane was 3 N m^{-2} , that of a typical conventional unmanned aerial vehicle is of the order of 10 N m^{-2} , and that of a modern civil airliner is of the order of $1,000 \text{ N m}^{-2}$.

Further improvements in these two performance limitations of overall efficiency and thrust density could enable EAD propulsion to open up new design spaces and unexplored applications for near-silent electric aircraft based on solid-state propulsion, in which the traditional limitations of propellers and gas turbines would no longer apply. Applications may include near-silent urban drones overcoming the current challenges of propeller noise, robust high-altitude environmental monitoring aeroplanes with no moving parts, and swarms of highly miniaturized solid-state-propulsion-based aircraft. The flight distances of 55 m and durations of 12 s for the heavier-than-air aircraft with solid-state propulsion described here compare well with the powered flight of the first heavier-than-air aircraft propelled by moving surfaces at Kitty Hawk, North Carolina, USA, 114 years ago. The Wright brothers achieved a flight¹⁸ of 35 m lasting 11 s—although they did have to carry a pilot rather than a remote control unit.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0707-9>.

Received: 22 June 2018; Accepted: 9 October 2018;

Published online 21 November 2018.

1. Chattock, A. P., Walker, W. E. & Dixon, E. H. IV. On the specific velocities of ions in the discharge from points. *Phil. Mag.* **1**, 79–98 (1901).
2. Stuetzer, O. M. Ion-drag pumps. *J. Appl. Phys.* **31**, 136–146 (1960).
3. Christenson, E. A. & Moller, P. S. Ion-neutral propulsion in atmospheric media. *AIAA J.* **5**, 1768–1773 (1967).
4. Wilson, J., Perkins, H. D. & Thompson, W. K. *An Investigation Of Ionic Wind Propulsion*. Report No. NASA/TM 2009–215822 (NASA, 2009).

5. Masuyama, K. & Barrett, S. R. H. On the performance of electrohydrodynamic propulsion. *Proc. R. Soc. A* **469**, 20120623 (2013).
6. Monroli, N., Ploouraboué, F. & Praud, O. Electrohydrodynamic thrust for in-atmosphere propulsion. *AIAA J.* **55**, 4296–4305 (2017).
7. Gilmore, C. K. & Barrett, S. R. H. Electrohydrodynamic thrust density using positive corona-induced ionic winds for in-atmosphere propulsion. *Proc. R. Soc. A* **471**, 20140912 (2015).
8. Loeb, L. B. *Electrical Coronas: Their Basic Physical Mechanisms* (Univ. California Press, Berkeley, 1965).
9. Melcher, J. R. Traveling-wave induced electroconvection. *Phys. Fluids* **9**, 1548 (1966).
10. Allen, P. H. G. & Karayiannis, T. G. Electrohydrodynamic enhancement of heat transfer and fluid flow. *Heat Recovery Syst.* **15**, 389–423 (1995).
11. Drew, D. S., Lambert, N. O., Schindler, C. B. & Pister, K. S. J. Toward controlled flight of the ionocraft: a flying microrobot using electrohydrodynamic thrust with onboard sensing and no moving parts. *IEEE Robotics Automation Lett.* **3**, 2807–2813 (2018).
12. Cumpsty, N. & Heyes, A. *Jet Propulsion* (Cambridge Univ. Press, Cambridge, 1998).
13. Leishman, J. G. *Principles of Helicopter Aerodynamics* (Cambridge Univ. Press, Cambridge, 2000).
14. Hoburg, W. & Abbeel, P. Geometric programming for aircraft design optimization. *AIAA J.* **52**, 2414–2426 (2014).
15. Gu, W. J. & Liu, R. A study of volume and weight vs. frequency for high-frequency transformers. In *Power Electronics Specialists Conf.* 1123–1129 (IEEE, 1993).
16. He, Y., Woolston, M. R. & Perreault, D. J. Design and implementation of a lightweight high-voltage power converter for electro-aerodynamic propulsion. In *IEEE Workshop on Control and Modeling for Power Electronics* <http://doi.org/10.1109/COMPEL.2017.8013315> (IEEE, 2017).
17. Hsu, W. C., Chen, J. F., Hsieh, Y. P. & Wu, Y. M. Design and steady-state analysis of parallel resonant DC–DC converter for high-voltage power generator. *IEEE Trans. Power Electronics* **32**, 957–966 (2017).
18. McFarland, M. W. (ed.) *The Papers of Wilbur and Orville Wright* (McGraw-Hill, New York, 1953).

Acknowledgements The work was also contributed to by many undergraduate students from 2010–2018 as part of MIT's Undergraduate Research Opportunities Program (UROP), as part of MIT's Minority Students Research Program (MSRP), or as part of MIT's summer research exchange program with Imperial College London (IROP). These students include Y. K. Tey, P. Kandangwa, W. B. Rideout, J. Epps, S. O'Neill, M. Adams, J. M. Salinas, N. H. Rodman, I. L. LaJoie, W. A. Rutter, A. J. Sanders, N. J. Martorell, I. Vallina Garcia, J. P. Liguori, K. Dasadikari, B. J. Scalzo Dees, M. H. Knowles, D. W. Fellows and D. P. Aaradhya. In addition, we thank A. Brown, T. Tao, C. Tan, P. Lozano, J. Peraire and C. Guerra-Garcia for technical discussions and advice, in some cases as part of student thesis committees. K. Masuyama, A. Dexter and J. Payton contributed to the project in its earlier phases. J. Leith and J. L. Freeman contributed to the financial and procurement administration for the project. F. Allroggen contributed to the resource management for the project. We also thank the laboratory staff at MIT AeroAstro for their help with the design, fabrication and flight testing of the EAD aircraft, in particular D. Robertson, T. Billings, A. Zolnik and T. Numan. Finally, we thank the MIT Department of Athletics, Physical Education, and Recreation for access to space for indoor flight testing, in particular S. Lett. This work was funded through MIT Lincoln Laboratory Autonomous Systems Line, the Professor Amar G. Bose Research Grant, and through the Singapore-MIT Alliance for Research and Technology (SMART). The work was also funded through the Charles Stark Draper and Leonardo career development chairs at MIT. This material is based on work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering.

Reviewer information Nature thanks D. Drew, K. Pister, F. Ploouraboué and H. Smith for their contribution to the peer review of this work.

Author contributions S.R.H.B. conceived the aeroplane. H.X. and C.K.G. designed the aeroplane. Y.H., D.J.P. and M.R.W. developed the electrical power systems. H.X., Y.H., K.L.S., C.K.G., S.P.K. and C.C.H. built and tested the aeroplane. H.X. piloted the aeroplane. K.L.S. and S.P.K. performed wind tunnel tests. S.R.H.B., D.J.P. and T.S. coordinated the project.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0707-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0707-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to S.R.H.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Aircraft design optimization. The geometric programming optimization, used to produce a minimum-wingspan aircraft, was implemented using the open source Python library GPykit. Geometric programming is a convex optimization method able to efficiently find the global optimum for a certain objective function given a series of constraints¹⁴. Geometric programming constraints are formulated as monomial or posynomial inequalities, which under a log transformation, are convex functions. The constraints used in our model were composed of flight equilibrium constraints, and drag, weight, structural, propulsion system and electrical models. As an example, the flight equilibrium constraint requires lift to be equal to the weight of the aircraft multiplied by some load factor n , which is 1 for steady-level flight:

$$nW \leq \frac{1}{2} \rho v_0^2 S C_L$$

where W is the weight of the aircraft, v_0 is the aircraft speed, S is the wing planform area, and C_L is the vehicle lift coefficient.

Some constraints were implemented through the fit of surrogate geometric programming compatible functions to outputs from higher-fidelity models. For example, the wing profile drag and lift coefficients were calculated for a range of Reynolds numbers using the two-dimensional viscous panel program XFOIL. The XFOIL output was fitted to a geometric-programming-compatible posynomial function, and used as part of the optimization. A total of 93 constraints were used in the GPykit model to solve for a set of 90 free design variables.

Electrical systems. The electrical system consisted of a battery pack and an HVPC. The battery was composed of 54 3.7-V E-Flite 150 mAh Li-Po (lithium-ion polymer) cells connected in series. This cell was chosen for its high discharge rate (45 C), which allowed a peak discharge current of 6.75 A. The nominal operating voltage of the battery stack was 200 V, but could vary between 150 V and 225 V depending on the state of charge and discharge rate. The battery could sustain a 600 W discharge power for approximately 90 s. The duration of the flight tests was limited by the length of the indoor space, rather than by battery endurance.

The custom-designed HVPC used a conventional topology consisting of an inverter, a high-voltage transformer and a Cockcroft–Walton rectifier (Fig. 1c). The inverter converts the battery pack voltage to a 500–700 kHz alternating current with a voltage gain of approximately 2.5. The high-voltage transformer steps this up with a 1:15 voltage gain to an approximately 7-kV alternating current. The Cockcroft–Walton multiplier rectifies the 7-kV alternating current and multiplies it with a voltage gain of 5.6, producing a direct-current voltage of up to 40.3 kV. The efficiency of the HVPC was 82%–85% depending on the output power draw.

Laboratory and wind tunnel tests. The static performance of the thruster was evaluated by hanging the aeroplane vertically from a laboratory scale and measuring the change in measured force when the thruster was turned on. This was used to estimate the thrust of the aeroplane in flight.

The dynamic thrust performance of a section of thruster was measured in a 0.48 m by 0.48 m open jet wind tunnel at 0 m s^{−1}, 2 m s^{−1} and 5 m s^{−1} using a Pitot tube wake momentum survey. The testing confirmed that EAD thrust remained constant at 1 N m^{−1} span as velocity increased, in agreement with theoretical predictions for low speeds, but the net force decreased as velocity increased owing to an increase in parasitic drag on the electrodes.

Flight tests. We conducted flight tests in the Johnson Athletic Track at MIT. The indoor space was chosen so that the aircraft could be operated in a controlled environment, minimizing the effects of wind and temperature. The space had a useable length of 60 m.

During powered flights, the aeroplane was launched and maintained on a straight and steady flight path. The rudder and elevator surfaces were actuated with electronic servos: they were trimmed for straight flight, but adjusted during flight by the pilot via remote radio control, in particular during launch and landing. Around 5–10 m from the end of the flight area, the HVPC was remotely powered off, and the aircraft glided to ground (Supplementary Video 1). Extended Data Fig. 2 shows the HVPC output voltage and input power during a flight.

Not all flights resulted in the same height or energy gain. Variation was introduced by (1) piloting inconsistencies in the application of rudder and elevator control, which increased energy dissipated by aircraft drag, and (2) mechanical damage to the electrodes between flights. Crashed landings at the ends of flights 1, 2 and 5, which caused physical damage to the electrodes, and their subsequent field repairs, resulted in varying thrust performance.

During unpowered glides, the propulsion system was not turned on (Supplementary Video 2). The plane was trimmed appropriately for the unpowered condition, and after launch, the pilot attempted to maintain the aeroplane at a nominal angle of attack. The loss of height was due to drag alone and not improper trim or pitch control.

Aircraft video tracking. To demonstrate that the aircraft had achieved steady-level flight, we needed accurate measurements of both the height above the ground (for gravitational potential energy), and the velocity (for kinetic energy). We developed a motion capture system for tracking of the aircraft over the 60 m flight space using four GoPro Hero 5 Black cameras operating at 60 frames per second and a resolution of 2,704 × 1,520 pixels. Three of these cameras were directed perpendicular to the flight path to track the aircraft position in the x – z (fore–aft and up–down) plane. The fourth was directed in the direction of flight to track in the y (spanwise) direction.

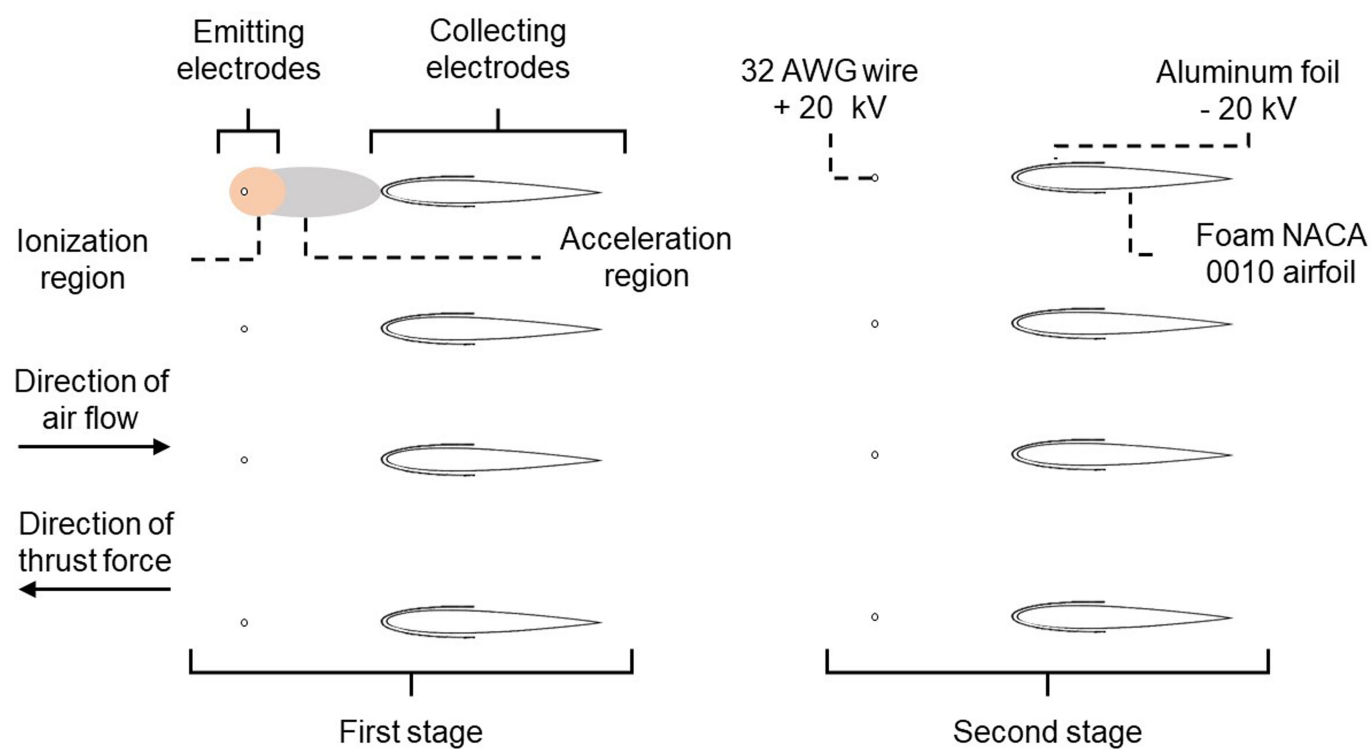
Two LED lights, one green and one red, were placed on the pylon and the tail of the aircraft, respectively, to allow triangulation of the centre of gravity and calculation of the pitch angle. These lights were located on each image frame using MATLAB's Computer Vision toolbox. The computer vision tracking error was estimated by the average radius of these coloured regions, which was 0.05–0.07 m.

The raw output images from the cameras were subject to lens distortion, which was corrected using MATLAB's Image Processing and Computer Vision toolbox. A set reference images were produced using physical markers spaced 5 m by 3 m by 1 m in the x , y and z directions. There was a root-mean-square error of 0.045 m between a reference point position estimated by camera tracking, and the real-world position of the reference point.

The error in position tracking was estimated in two ways. The first was calculated as the sum of the computer vision tracking error, which was estimated from the size of the colored regions, and the lens correction error, which was estimated by the root-mean-square error of the reference point positions. The second was calculated from the regions of overlap between the cameras, where the plane was simultaneously in the field of view of more than one camera. By comparing the position reported by one camera with that simultaneously reported by another, a second estimate of each camera's tracking error was obtained. These errors were of similar magnitude, and the greater of the two errors was used to estimate the overall tracking error. This was 0.10–0.15 m depending on the particular flight, and this is the error shown in Figs. 3 and 4.

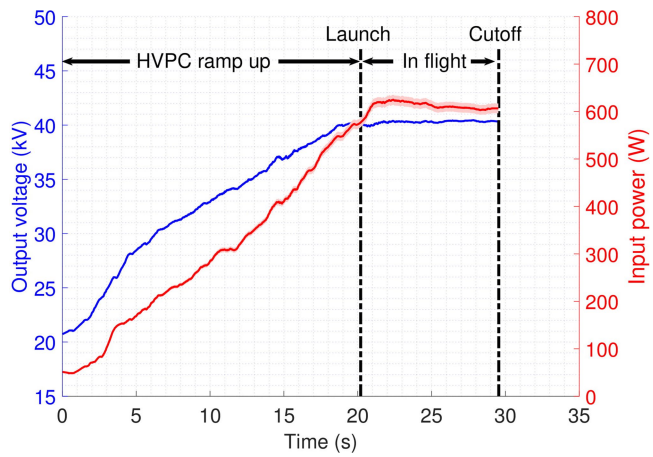
Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.



Extended Data Fig. 1 | Schematic of propulsion system electrodes. Not to scale. The emitting electrode is a 32 American Wire Gauge (32 AWG; 0.2 mm diameter) stainless steel wire, held at 60 mm spacing from the collecting electrode by 3D-printed spacers. The collecting electrode is a

National Advisory Committee for Aeronautics (NACA) 0010 airfoiled foam section covered in a thin layer of aluminium foil. The electrodes are 3 m in span (into the page).



Extended Data Fig. 2 | HVPC output voltage and input power for a single flight (number 9). The HVPC is designed to ramp up to the final voltage over 20 s while the aeroplane is on the launcher. The aircraft was in flight for 10–12 s. During flight, the HVPC regulates the output voltage to maintain 40.3 kV.

Extended Data Table 1 | Key engineering parameters and performance metrics of the EAD aeroplane

Mass Budget	Total (kg)	2.45
	Power converter (kg)	0.51
	Battery (kg)	0.23
	Wing (kg)	0.63
	Electrodes (kg)	0.41
Aerodynamic Characteristics	Wing Span (m)	5.14
	Flight Velocity (m/s)	4.8 ± 0.2
	Aspect Ratio	17.9
	Drag (N)	3.0 ± 0.2
	Lift/Drag Ratio	8 ± 1
EAD Propulsion System	Thrust (N)	3.2 ± 0.2
	Voltage (kV)	40.3 ± 0.1
	Power Requirement (W)	620 ± 20
	Thrust Frontal Area (m ²)	0.9

Thrust and drag values are estimated from flight test data, and laboratory and wind tunnel experiments. Aerodynamic performance uncertainties arise from flight tracking error (see Methods). Uncertainties in electrical data are the 95% confidence interval from onboard HVPC measurements.

Efficient radical-based light-emitting diodes with doublet emission

Xin Ai^{1,3}, Emrys W. Evans^{2,3}, Shengzhi Dong^{1,3}, Alexander J. Gillett², Haoqing Guo¹, Yingxin Chen¹, Timothy J. H. Hele², Richard H. Friend^{2*} & Feng Li^{1,2*}

Organic light-emitting diodes (OLEDs)^{1–5}, quantum-dot-based LEDs^{6–10}, perovskite-based LEDs^{11–13} and micro-LEDs^{14,15} have been championed to fabricate lightweight and flexible units for next-generation displays and active lighting. Although there are already some high-end commercial products based on OLEDs, costs must decrease whilst maintaining high operational efficiencies for the technology to realise wider impact. Here we demonstrate efficient action of radical-based OLEDs¹⁶, whose emission originates from a spin doublet, rather than a singlet or triplet exciton. While the emission process is still spin-allowed in these OLEDs, the efficiency limitations imposed by triplet excitons are circumvented for doublets. Using a luminescent radical emitter, we demonstrate an OLED with maximum external quantum efficiency of 27 per cent at a wavelength of 710 nanometres—the highest reported value for deep-red and infrared LEDs. For a standard closed-shell organic semiconductor, holes and electrons occupy the highest occupied and lowest unoccupied molecular orbitals (HOMOs and LUMOs), respectively, and recombine to form singlet or triplet excitons. Radical emitters have a singly occupied molecular orbital (SOMO) in the ground state, giving an overall spin-1/2 doublet. If—as expected on energetic grounds—both electrons and holes occupy this SOMO level, recombination returns the system to the ground state, giving no light emission. However, in our very efficient OLEDs, we achieve selective hole injection into the HOMO and electron injection to the SOMO to form the fluorescent doublet excited state with near-unity internal quantum efficiency.

For the most part, stable, organic luminescent radicals have been considered as curiosities with limited applications.^{17–23} Photoexcitation of doublet-ground-state (D_0) molecules generates doublet excited states, and spin-allowed emission—that is, fluorescence—in these molecules originates from the lowest-lying doublet excited state, D_1 (Fig. 1a).

By incorporating 3-substituted-9-(naphthalen-2-yl)-9H-carbazole (3NCz) and 3-substituted-9-phenyl-9H-carbazole (3PCz) to the core tris(2,4,6-trichlorophenyl)methyl (TTM) radical, we obtained two new luminescent radicals, TTM-3NCz and TTM-3PCz (Fig. 1b). The photoluminescence quantum efficiency (PLQE) in solid 4,4-bis(carbazol-9-yl)biphenyl (CBP) matrix film (3.0 wt%) is $(85.6 \pm 5.4)\%$ and $(60.4 \pm 0.9)\%$ for deep-red emission in TTM-3NCz (707 nm) and TTM-3PCz (695 nm), respectively, which can be translated to excellent device performance. See dashed lines in Fig. 1c for the photoluminescence spectra (photoexcitation at 375 nm).

A series of OLEDs using TTM-3NCz and TTM-3PCz as emitters were fabricated by vacuum deposition processing (pressure $< 6 \times 10^{-7}$ torr). The evaporation temperatures of TTM-3NCz and TTM-3PCz under vacuum are below 473 K, much lower than their respective thermal decomposition temperatures of 635 K and 640 K (Extended Data Fig. 1a), meaning that their thermal stabilities are sufficient to withstand the thermal-evaporation process. The energy levels of the two compounds were obtained from cyclic voltammetry measurements (Extended Data Fig. 2a, c). Furthermore, to assess

the electrochemical stabilities of TTM-3NCz and TTM-3PCz, we obtained cyclic voltammetry curves over 20 scanning cycles (Extended Data Fig. 2b, d). There are no substantial changes between the curves, which indicates good redox stability for TTM-3NCz and TTM-3PCz (in addition to good photostability; Extended Data Fig. 3).

1-Bis[4-[N,N-di(4-tolyl)amino]phenyl]cyclohexane (TAPC)¹ and 2,4,6-tris[m-(diphenylphosphinoyl)phenyl]-1,3,5-triazine (PO-T2T)²⁴ were used to fabricate 35-nm-thick hole- and 70-nm-thick electron-transport layers, respectively. The radicals were doped into a CBP host to form the light-emitting layer (thickness, 25–40 nm). Owing to the larger energy bandgap of CBP compared to that of the dopant molecules (see Fig. 1d inset), sequential charge trapping is expected to be the main route for the creation of doublet excitons. A thin layer (10 nm) of 4,6-Bis(3,5-di(pyridin-3-yl)phenyl)-2-methylpyrimidine (B3PYMPM)²⁵ was inserted between the light-emitting and PO-T2T layers to remove unwanted green emission (probably CBP:PO-T2T exciplex). The best LED performance was found for ITO/MoO₃ (3 nm)/TAPC (35 nm)/CBP:TTM-3NCz (3.0 wt%; (40 nm) and CBP:TTM-3PCz (3.0 wt%; 25 nm)/B3PYMPM (10 nm)/PO-T2T (70 nm)/LiF (0.8 nm)/Al (100 nm).

Plots of the external quantum efficiency (EQE) against the current density for the OLEDs are given in Fig. 1d (TTM-3NCz, red; TTM-3PCz, black). The corresponding electroluminescence peaks appear at 710 nm (TTM-3NCz) and 703 nm (TTM-3PCz), and the devices have true deep-red emission (Fig. 1c). The maximum EQE values of $(27 \pm 5)\%$ for the TTM-3NCz OLEDs and $(17 \pm 3)\%$ for the TTM-3PCz OLEDs suggest internal quantum efficiency near 100% for electroluminescence, when considered with the film PLQE values and a 30% light outcoupling coefficient⁵. To the best of our knowledge, the maximum EQE of the TTM-3NCz-based device is the highest value reported so far for deep-red/infrared LEDs^{26–28} (see Supplementary Table 2), whereas respectable EQE values $> 10\%$ are obtained at 1 mA cm^{-2} . The near-identical electroluminescence and photoluminescence spectra (Fig. 1c) for TTM-3NCz and TTM-3PCz show that electroluminescence and photoluminescence emission originate from the same electronic transition ($D_1 \rightarrow D_0$).

The current density–voltage electroluminescence characteristics of the best TTM-3NCz- and TTM-3PCz-based devices are given in Fig. 1e, f. The data points associated with the maximum EQE values are denoted by arrows and occur in the device turn-on regions, above the experimental noise levels. The performance of five more TTM-3NCz ($\text{EQE}_{\text{max}} = 27\%–16\%$) devices is shown in Extended Data Fig. 4. The electroluminescence spectra are unchanged for a wide range of operating current densities, $6 \mu\text{A cm}^{-2}$ – 1.6 mA cm^{-2} (Extended Data Fig. 5).

We note that TTM-3NCz and TTM-3PCz contain ‘donor’ 3NCz/3PCz and ‘acceptor’ TTM radical groups, which resemble the structural motif of classical thermally activated delayed fluorescence (TADF) molecules. In Fig. 2a absorption and photoluminescence spectra for the molecules are plotted with reference to TTM. Introduction of the ‘donor’ group to the TTM moiety leads to the appearance of a new

¹State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun, China. ²Cavendish Laboratory, University of Cambridge, Cambridge, UK. ³These authors contributed equally: Xin Ai, Emrys W. Evans, Shengzhi Dong. *e-mail: rhf10@cam.ac.uk; lifeng01@jlu.edu.cn

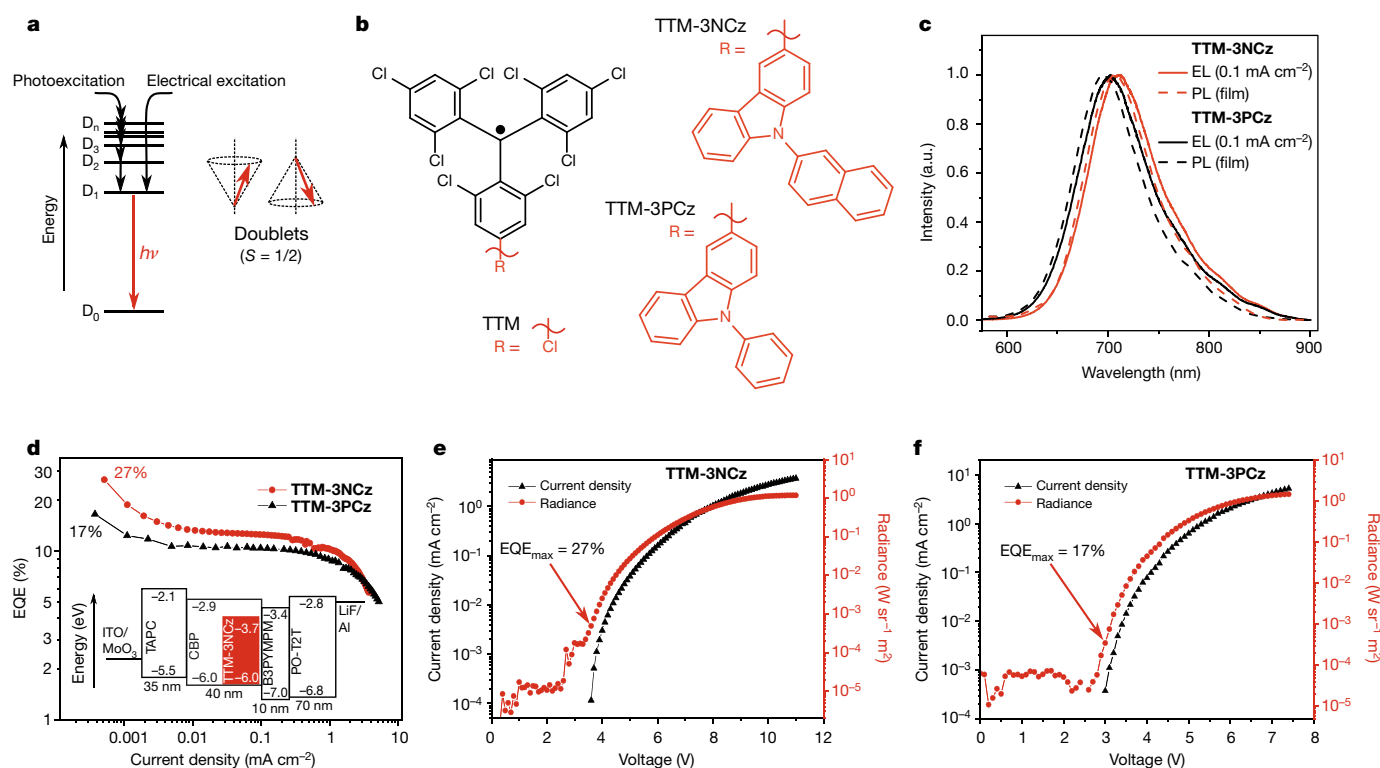


Fig. 1 | LEDs with doublet emission. **a**, Doublet emission following photo- and electrical excitation. The illustration on the right indicates the electron spin vector representation for doublets. **b**, Chemical structures of TTM, TTM-3NCz and TTM-3PCz. **c**, Electroluminescence (EL; solid lines) and photoluminescence (PL; dotted lines; photoexcitation wavelength, 375 nm) spectra for TTM-3NCz (red) and TTM-3PCz (black). **d**, EQE-current density curves for TTM-3NCz (red) and TTM-3PCz (black) LEDs. The inset shows the TTM-3NCz device layout; the

labels give the energy levels in electronvolts and the thickness of layers in nanometres. For the best TTM-3NCz LED architecture, SOMO = -3.7 eV (defined by cyclic voltammetry; see Extended Data Fig. 2) and HOMO = -6.0 eV (approximated by 9-phenylcarbazole). **e**, **f**, Black triangles and red circles denote, respectively, current density-voltage and radiance-voltage profiles for TTM-3NCz (**e**) and TTM-3PCz (**f**). Radiance levels corresponding to EQE_{max} lie above the background noise level.

absorption band at about 620 nm and an accompanying red shift of about 0.38 eV in photoluminescence. A strong charge-transfer character is expected for the first excited state of TTM-3NCz and TTM-3PCz; that is, substantial spatial separation with little overlap for the 3NCz/3PCz-centred HOMO and TTM-centred SOMO. These frontier molecular orbitals can be used to describe electronic transitions for ground-state (D_0) absorption to the lowest excited state (D_1) and in photo- and electroluminescence ($D_1 \rightarrow D_0$). The dipole moment, δ , of D_1 is expected to have the same orientation as (δ^-) TTM-(δ^+) donor in D_0 , but with greater magnitude. This is supported by the observation of strong, positive solvatochromic effects (see Fig. 2a, b; photoexcitation at 375 nm). Increasing the polarizability index of the solvent leads to increasing Stokes shifts. The slope of the Lippert–Mataga plot of the TTM-3NCz molecules in Fig. 2c reflects the change in dipole moment ($\Delta\delta$) upon photoexcitation, in contrast to the solvent-independent behaviour (that is, $\Delta\delta \approx 0$) of TTM. In TTM the D_1 excited state is more locally excited in nature (that is, it has considerable HOMO/SOMO overlap).

For good LED performance it is critical that the dopant emitters have high PLQE. PLQE values of 49% and 46% are obtained for toluene solutions of TTM-3NCz and TTM-3PCz, respectively. The energy gap law generally precludes efficient deep-red/infrared light emission, but does not appear to be strictly followed by dopants with appreciable charge-transfer character in emission²⁷. In favour of the examined OLEDs, the non-radiative decay pathways were found to be further reduced in CBP films doped at 3.0 wt%, giving rise to the PLQE values of $(85.6 \pm 5.4)\%$ (TTM-3NCz) and $(60.4 \pm 0.9)\%$ (TTM-3PCz) reported above.

To explore the nature of the doublet excited states, we performed nanosecond transient-absorption and -photoluminescence studies on toluene solutions containing TTM-3NCz and TTM-3PCz. Excitation

wavelengths of 532 nm and 600 nm were used for transient absorption and photoluminescence, respectively, chosen to excite the broad absorption band associated with the $D_0 \rightarrow D_1$ transition. In the measurements of nanosecond transient photoluminescence, we observe an emission spectrum that closely resembles the steady-state photoluminescence for TTM-3NCz and TTM-3PCz. There is no temporal evolution in the spectra. Furthermore, from the overlaid transient-absorption and -photoluminescence profiles (Fig. 2d), it is apparent that there are no excited-state species surviving beyond the emission lifetime. Although dark, triplet states usually occur at lower energy than emissive singlet states for closed-shell systems, we consider that there are no equivalent triplet states to hinder emission for the open-shell TTM-3NCz and TTM-3PCz molecules. Further discussion about the transient absorption measurements can be found in Supplementary Information, Section 2.

The photoluminescence kinetics is fitted with a mono-exponential function and yields lifetimes of 17.2 ns (TTM-3NCz) and 21.2 ns (TTM-3PCz). By combining these lifetimes with the PLQE values, the radiative decay rates are $2.9 \times 10^7 \text{ s}^{-1}$ (TTM-3NCz) and $2.1 \times 10^7 \text{ s}^{-1}$ (TTM-3PCz). These values are an order of magnitude larger than those associated with delayed emission from typical TADF molecules (for example, 2,4,5,6-tetra(9H-carbazol-9-yl)isophthalonitrile, 4CzIPN). This is particularly important for LEDs because higher radiative rates mean reduced exciton-charge annihilation issues with increasing current density.

Finally, the molecular properties and photophysics of the TTM-3NCz and TTM-3Cz molecules can be combined with density functional theory (DFT) calculations. Using unrestricted Kohn–Sham (UKS) DFT and time-dependent DFT (TDDFT) calculations with a B3LYP functional and a 6-31G** basis set, the nature of the electronic states is revealed. The interpretation of the computational results is shown

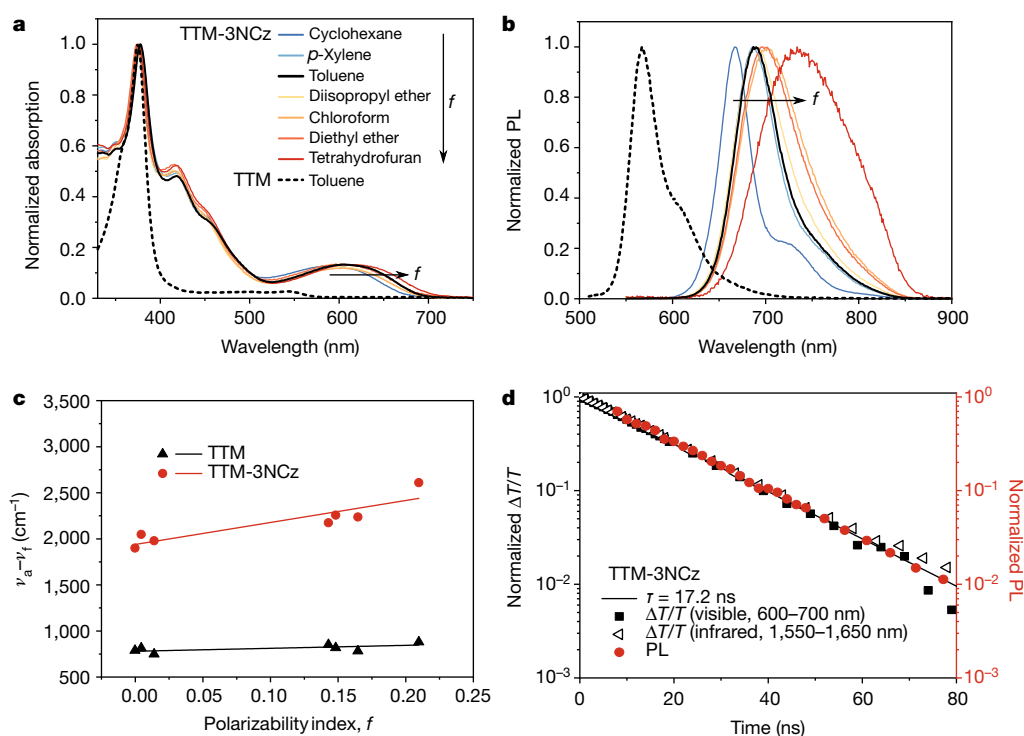


Fig. 2 | Doublet photophysics. **a**, **b**, Steady-state ultraviolet–visible (**a**) and photoluminescence (**b**) profiles for 10 μM TTM-3NCz in solvents of varying polarizability index, f . Reference measurements for 10 μM TTM in toluene are denoted by the black dotted line. Photoexcitation wavelength, 375 nm. **c**, Lippert–Mataga plot of the Stokes shift ($\nu_a - \nu_f$) versus f for TTM-3NCz (red circles) and TTM (black triangles). ν_a and ν_f

denote the absorption and fluorescence energies, respectively. **d**, Kinetic profiles for photoluminescence (integrated in 650–850 nm), red circles; 600 nm excitation at 6.5 $\mu\text{J cm}^{-2}$ fluence) and transient absorption (600–700 nm averaged, black squares; 1,550–1,650 nm averaged, white triangles; 532 nm excitation at 35.7 $\mu\text{J cm}^{-2}$ fluence). The solid black line shows the mono-exponential fit. The sample concentration is 10–100 μM .

in the molecular orbital diagram of Fig. 3a. The scheme begins with TTM and considers interactions between benzene HOMO/LUMO moieties and the central carbon $2p_z$ orbital. A relatively strong absorption

peak at 374–378 nm found for TTM and TTM–donor-type molecules is attributed to a SOMO \rightarrow LUMO transition. The SOMO has electron density on every other atom of the TTM group and can be determined

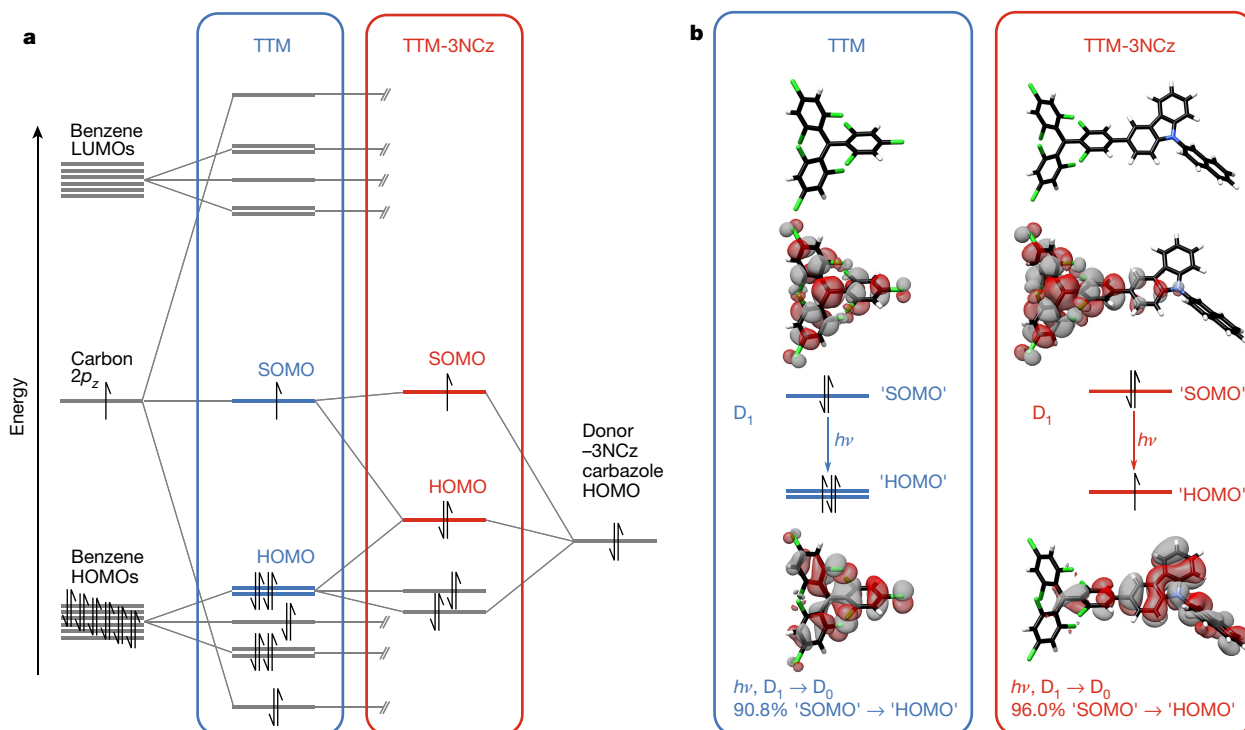


Fig. 3 | Electronic structure of doublets: from HOMO to SOMO. **a**, Molecular orbital diagrams for TTM and TTM-3NCz. Structures are geometry-optimized for the D_1 state using UKS-TDDFT (B3LYP, 6-31G**). The half-arrows indicate the state occupancy and electron

spin orientation. **b**, Molecular orbitals involved in the mono-electronic depiction for $D_1 \rightarrow D_0 + h\nu$. The labels 'HOMO' and 'SOMO' refer to charged HOMO and SOMO orbitals.

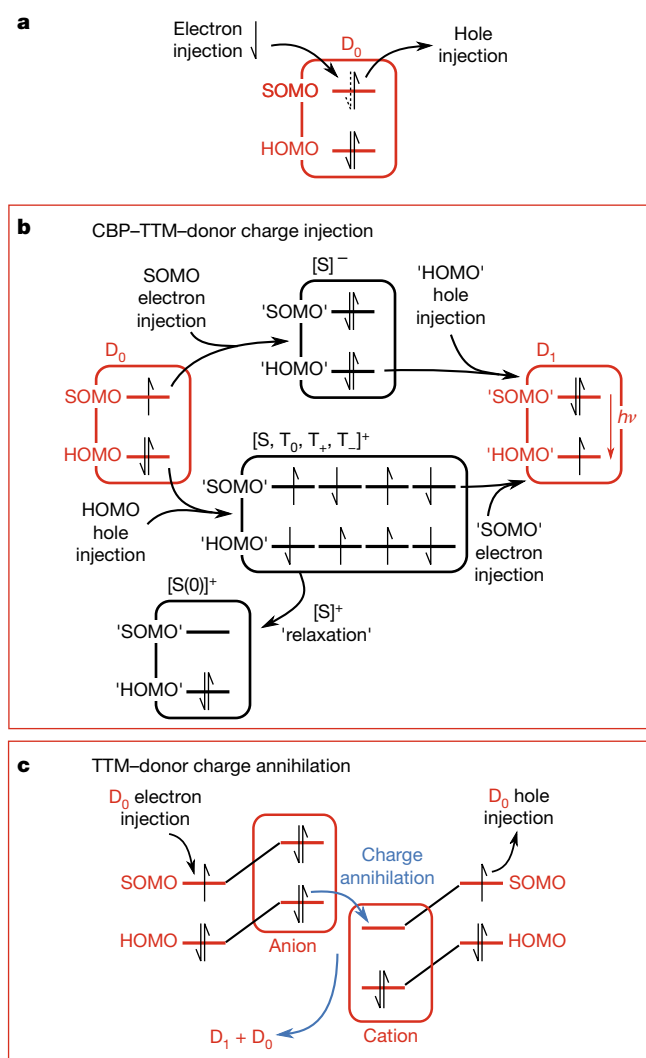


Fig. 4 | Doublet electroluminescence mechanism. **a**, Schematic depicting how thermodynamically favoured SOMO electron and hole injection does not realize the emissive doublet excited state. The dashed arrow indicates a hole. **b**, Electroluminescence by selective HOMO and SOMO hole and electron injection. The scheme shows two routes to creating D_1 : first electron injection, then hole injection (top); first hole injection, then electron injection (bottom). $[S]^-$, $[S]^+$, $[S(0)]^+$ and $[S, T_0, T_+, T_-]^+$ denote positively and negatively charged intermediates, with the electron occupancy depicted in the figure. 'HOMO' and 'SOMO' refer to charged HOMO and SOMO orbitals. **c**, Electroluminescence generated by TTM-3NCz/TTM-3PCz charge annihilation. This mechanism considers that the energy levels of TTM-donor-type molecules change following electron and hole injection to favour D_1 formation.

from first principles, along with the overall HOMO, using the group and Hückel theories (as outlined in Supplementary Information Section 3). The derived SOMO has the same form as that obtained from DFT (Fig. 3b). In TTM, TTM-3NCz and TTM-3PCz, the overall LUMO mirrors the LUMO of the -TTM benzene groups (Fig. 3a).

HOMO \rightarrow SOMO transitions give rise to the lowest energy absorption bands in both TTM-3NCz (616 nm) and TTM (541 nm) in toluene, with substantial charge-transfer and locally excited character, respectively. This arises because the TTM-3NCz and TTM-3PCz HOMOs are primarily hybrids of the most-anti-bonding combinations of TTM and carbazole-group HOMOs (Fig. 3b). Going from TTM to TTM-3NCz, the TTM molecular orbital diagram is perturbed, as depicted in Fig. 3a. Unexpectedly good agreement is found between the absorption spectra and the UKS-TDDFT²⁹ calculations for the HOMO \rightarrow SOMO peak positions in TTM-3NCz: experimental, 616 nm; calculated, 622 nm. A more detailed

discussion about the molecular orbital diagram and electronic structure calculations can be found in Supplementary Information Section 3 and 4.

The SOMO \rightarrow HOMO transitions, as depicted in Fig. 3b, are associated with luminescence. However, although the route to efficient doublet emission for open-shell molecules must involve these orbitals, thermodynamics dictates that electrons and holes are both stabilized by occupancy of the SOMO level following electrical injection (see Fig. 4a). Therefore, it is not possible to realize doublet excited states for light emission with this scheme.

As shown in the inset of Fig. 1d, the electrodes are biased for selective hole and electron injection into the 3NCz/3PCz HOMO and TTM SOMO levels, respectively (energy gap law). If injection of a hole into the HOMO occurs before that of an electron into the SOMO, positively charged singlet and triplet intermediates are expected (see Fig. 4b). It is possible that the positively charged singlet intermediate in the former case can 'relax' before the electron is injected into the SOMO, but this loss pathway (the relaxation of the positively charged singlet intermediate) to D_1 formation is spin-forbidden for the triplet intermediates. On the other hand, if injection of the electron into the SOMO occurs first, this gives a negatively charged singlet intermediate ($[S]^-$; Fig. 4b). In this second scenario, holes that travel via CBP HOMO levels could encounter negatively charged TTM-3NCz sites to form D_1 , by tunnelling between the carbazoles of CBP and TTM-3NCz. The holes could also hop between TTM-3NCz sites when avoiding the aforementioned singlet 'relaxation'. It is noteworthy that TTM-donor-type molecules have been found to make much better OLEDs than TTM³⁰, which is indirect evidence supporting the mechanism shown in Fig. 4b.

The scheme in Fig. 4b considers hole injection sequentially into the CBP host and then into the TTM-3NCz HOMO level. We can also consider models that do not include the CBP, in which charge annihilation of TTM-3NCz anions and cations generates the same D_1 excited state. This would require energy levels to be substantially lowered and raised following oxidation and reduction, respectively, so that the SOMO and HOMO of TTM-3NCz would become more favourably aligned for electron transfer to yield $D_0 + D_1$, as illustrated in Fig. 4c. Although energy shifts are expected upon charging, cyclic voltammetry data (Extended Data Fig. 2) suggest that these shifts are too small to achieve the energy level alignment indicated in Fig. 4c. Therefore at this stage, the precise route from electrical injection to luminescence is unclear but undoubtedly efficient.

We have demonstrated highly efficient radical-based OLEDs with EQE values that exceed by far those of other LEDs with deep-red/near-infrared emission. Our scheme is based on using open-shell, doublet dopants that emit light after donor–radical charge transfer. The SOMO in these molecules facilitates the exceptional performance of the LEDs and its spin properties offer many possibilities for application in other fields of optoelectronics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0695-9>.

Received: 12 June; Accepted: 2 October 2018;

Published online 21 November 2018.

1. Tang, C. W. & VanSlyke, S. A. Organic electroluminescent diodes. *Appl. Phys. Lett.* **51**, 913–915 (1987).
2. Burroughes, J. H. et al. Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990); correction **348**, 352 (1990).
3. Baldo, M. A. et al. Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395**, 151–154 (1998).
4. Ma, Y., Zhang, H., Shen, J. & Che, C. Electroluminescence from triplet metal-ligand charge-transfer excited state of transition metal complexes. *Synth. Met.* **94**, 245–248 (1998).
5. Uoyama, H., Goushi, K., Shizu, K., Nomura, H. & Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **492**, 234–238 (2012).

6. Tessler, N., Medvedev, V., Kazes, M., Kan, S. & Banin, U. Efficient near-infrared polymer nanocrystal light-emitting diodes. *Science* **295**, 1506–1508 (2002).
7. Sun, Q. et al. Bright, multicoloured light-emitting diodes based on quantum dots. *Nat. Photon.* **1**, 717–722 (2007).
8. Dai, X. et al. Solution-processed, high-performance light-emitting diodes based on quantum dots. *Nature* **515**, 96–99 (2014).
9. Yang, Y. et al. High-efficiency light-emitting devices based on quantum dots with tailored nanostructures. *Nat. Photon.* **9**, 259–266 (2015).
10. Dai, X., Deng, Y., Peng, X. & Jin, Y. Quantum-dot light-emitting diodes for large-area displays: towards the dawn of commercialization. *Adv. Mater.* **29**, 1607022 (2017).
11. Tan, Z.-K. et al. Bright light-emitting diodes based on organometal halide perovskite. *Nat. Nanotechnol.* **9**, 687–692 (2014).
12. Cho, H. et al. Overcoming the electroluminescence efficiency limitations of perovskite light-emitting diodes. *Science* **350**, 1222–1225 (2015).
13. Wang, N. et al. Perovskite light-emitting diodes based on solution-processed self-organized multiple quantum wells. *Nat. Photon.* **10**, 699–704 (2016).
14. Jin, S. X., Li, J., Li, J. Z., Lin, J. Y. & Jiang, H. X. GaN microdisk light emitting diodes. *Appl. Phys. Lett.* **76**, 631–633 (2000).
15. Zhang, K., Peng, D., Lau, K. M. & Liu, Z. Fully-integrated active matrix programmable UV and blue micro-LED display system-on-panel (SoP). *J. Soc. Inf. Disp.* **25**, 240–248 (2017).
16. Peng, Q., Obolda, A., Zhang, M. & Li, F. Organic light-emitting diodes using a neutral π radical as emitter: the emission from a doublet. *Angew. Chem. Int. Ed.* **54**, 7091–7095 (2015).
17. Ballester, M., Molinet, C. & Castañer, J. Preparation of highly strained aromatic chlorocarbons. I. A powerful nuclear chlorinating agent. Relevant reactivity phenomena traceable to molecular strain. *J. Am. Chem. Soc.* **82**, 4254–4258 (1960).
18. Armet, O. et al. Inert carbon free radicals. 8. Polychlorotriphenylmethyl radicals: synthesis, structure, and spin-density distribution. *J. Phys. Chem.* **91**, 5608–5616 (1987).
19. Heckmann, A., Lambert, C., Goebel, M. & Wortmann, R. Synthesis and photophysics of a neutral organic mixed-valence compound. *Angew. Chem. Int. Ed.* **43**, 5851–5856 (2004).
20. Velasco, D. et al. Red organic light-emitting radical adducts of carbazole and tris(2,4,6-trichlorotriphenyl)methyl radical that exhibit high thermal stability and electrochemical amphotericity. *J. Org. Chem.* **72**, 7523–7532 (2007).
21. Castellanos, S., Velasco, D., López-Calahorra, F., Brillas, E. & Julia, L. Taking advantage of the radical character of tris(2,4,6-trichlorophenyl)methyl to synthesize new paramagnetic glassy molecular materials. *J. Org. Chem.* **73**, 3759–3767 (2008).
22. Hattori, Y., Kusamoto, T. & Nishihara, H. Luminescence, stability, and proton response of an open-shell (3,5-dichloro-4-pyridyl)bis(2,4,6-trichlorophenyl)methyl radical. *Angew. Chem. Int. Ed.* **53**, 11845–11848 (2014).
23. Ai, X., Chen, Y., Feng, Y. & Li, F. A stable room-temperature luminescent biphenylmethyl radical. *Angew. Chem. Int. Ed.* **57**, 2869–2873 (2018).
24. Hung, W. Y. et al. The first tandem, all-exciplex-based WOLED. *Sci. Rep.* **4**, 5161 (2014).
25. Sasabe, H. et al. 2-Phenylpyrimidine skeleton-based electron-transport materials for extremely efficient green organic light-emitting devices. *Chem. Commun.* 5821–5823 (2008).
26. Li, C. et al. Deep-red to near-infrared thermally activated delayed fluorescence in organic solid films and electroluminescent devices. *Angew. Chem. Int. Ed.* **56**, 11525–11529 (2017).
27. Kim, D.-H. et al. High-efficiency electroluminescence and amplified spontaneous emission from a thermally activated delayed fluorescent near-infrared emitter. *Nat. Photon.* **12**, 98–104 (2018).
28. Xue, J. et al. High-efficiency near-infrared fluorescent organic light-emitting diodes with small efficiency roll-off: a combined design from emitters to devices. *Adv. Funct. Mater.* **27**, 1703283 (2017).
29. Ipatov, A. et al. Excited-state spin-contamination in time-dependent density-functional theory for molecules with open-shell ground states. *J. Mol. Struct. Theochem* **914**, 60–73 (2009).
30. Neier, E. et al. Solution-processed organic light-emitting diodes with emission from a doublet exciton; using (2,4,6-trichlorophenyl)methyl as emitter. *Org. Electron.* **44**, 126–131 (2017).

Acknowledgements X.A., S.D., H.G., Y.C. and F.L. are grateful for the financial support received from the National Key R&D Program of China (grant number 2016YFB0401001), the National Natural Science Foundation of China (grant numbers 51673080 and 91233113) and the National Key Basic Research and Development Program of China (973 programme, grant number 2015CB655003). E.W.E., A.J.G. and R.H.F. thank the EPSRC for funding (EP/M01083X/1, EP/M005143/1). T.J.H.H. thanks Jesus College, Cambridge for a Research Fellowship. F.L. is an academic visitor at the Cavendish Laboratory, Cambridge and is supported by the China Scholarship Council (CSC) and the Talents Cultivation Program (Jilin University, China).

Reviewer information *Nature* thanks T. Kusamoto and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions X.A., S.D. and H.G. designed and synthesized the luminescent radicals and performed the steady-state spectroscopy. E.W.E. performed the transient-photoluminescence measurements and the quantum chemical calculations. T.J.H.H. devised the group theory treatment. A.J.G. conducted the transient-absorption spectroscopy measurements. X.A., Y.C. and F.L. optimized the devices. E.W.E., R.H.F. and F.L. initiated, designed and supervised the work. E.W.E., R.H.F. and F.L. wrote the manuscript, with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0695-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0695-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to R.H.F. or F.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Synthesis of doublet emitters. The precursors of TTM-3NCz and TTM-3PCz were prepared by Suzuki coupling of tris(2,4,6-trichlorophenyl)methane (HTTM) and 4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl-3NCz/3PCz. Radicals were generated from the precursors by treatment with potassium *t*-butoxide in tetrahydrofuran, followed by oxidation of the resulting carbanions with *p*-chloranil. The full details of the synthesis and characterization are provided in Supplementary Information, Section 1.2.

Device physics. the current density–voltage–electroluminescence characteristics were measured using a Keithley 2400 source meter, a Keithley 2000 multimeter and a calibrated silicon photodiode.

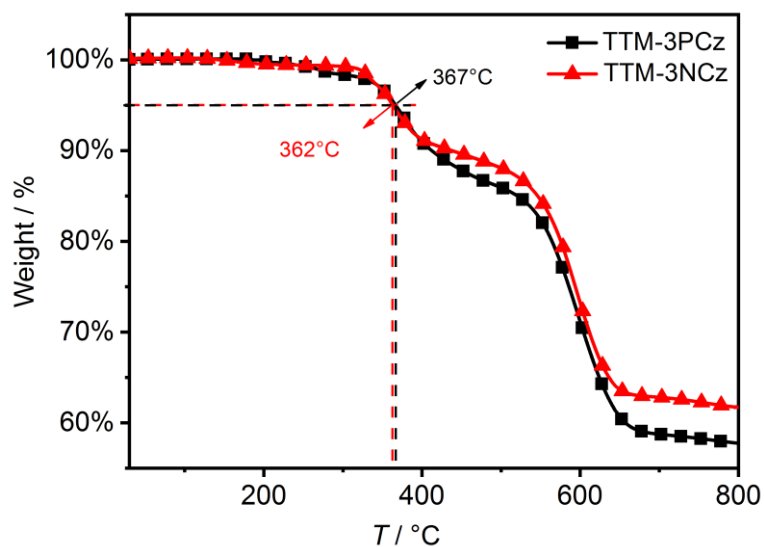
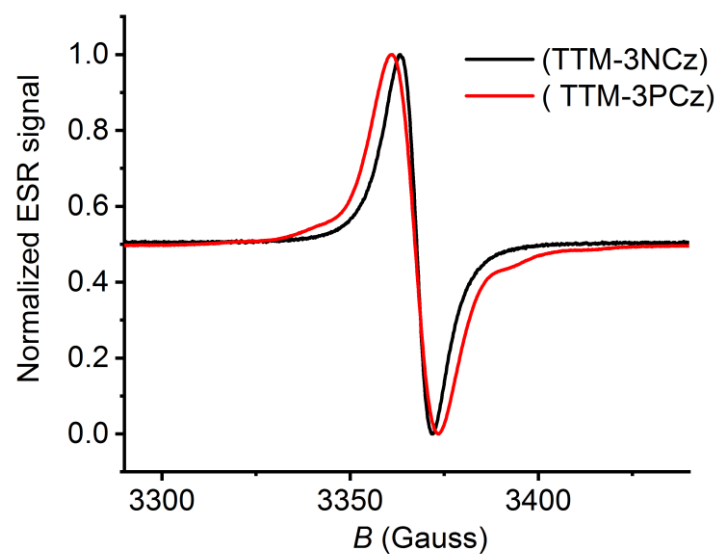
Photophysics. Ultraviolet/visible absorption spectra were measured with a Shimadzu UV-2550 spectrometer. Fluorescence spectra were recorded using a Shimadzu 5301PC spectrometer. The absolute fluorescence quantum yields

were obtained using an Edinburgh Instruments FLS920 spectrometer with the integrating-sphere method. Photoluminescence lifetimes were measured either by an Edinburgh Instruments FLS980 spectrometer (time resolution) or by an Andor iStar DH740 CCI-010 ICCD camera and an Andor SR303i spectrograph (time and spectral resolution). Experimental details for the transient absorption measurements are given in Supplementary Information.

Electronic structure calculations. For the ground state, UKS-DFT calculations were performed with the Orca package (version 4.0.1) using the B3LYP functional and the 6-31G** basis set. For the excited states, UKS-TDDFT was employed with the Tamm–Dancoff approximation. All calculations were treated in vacuo.

Data availability

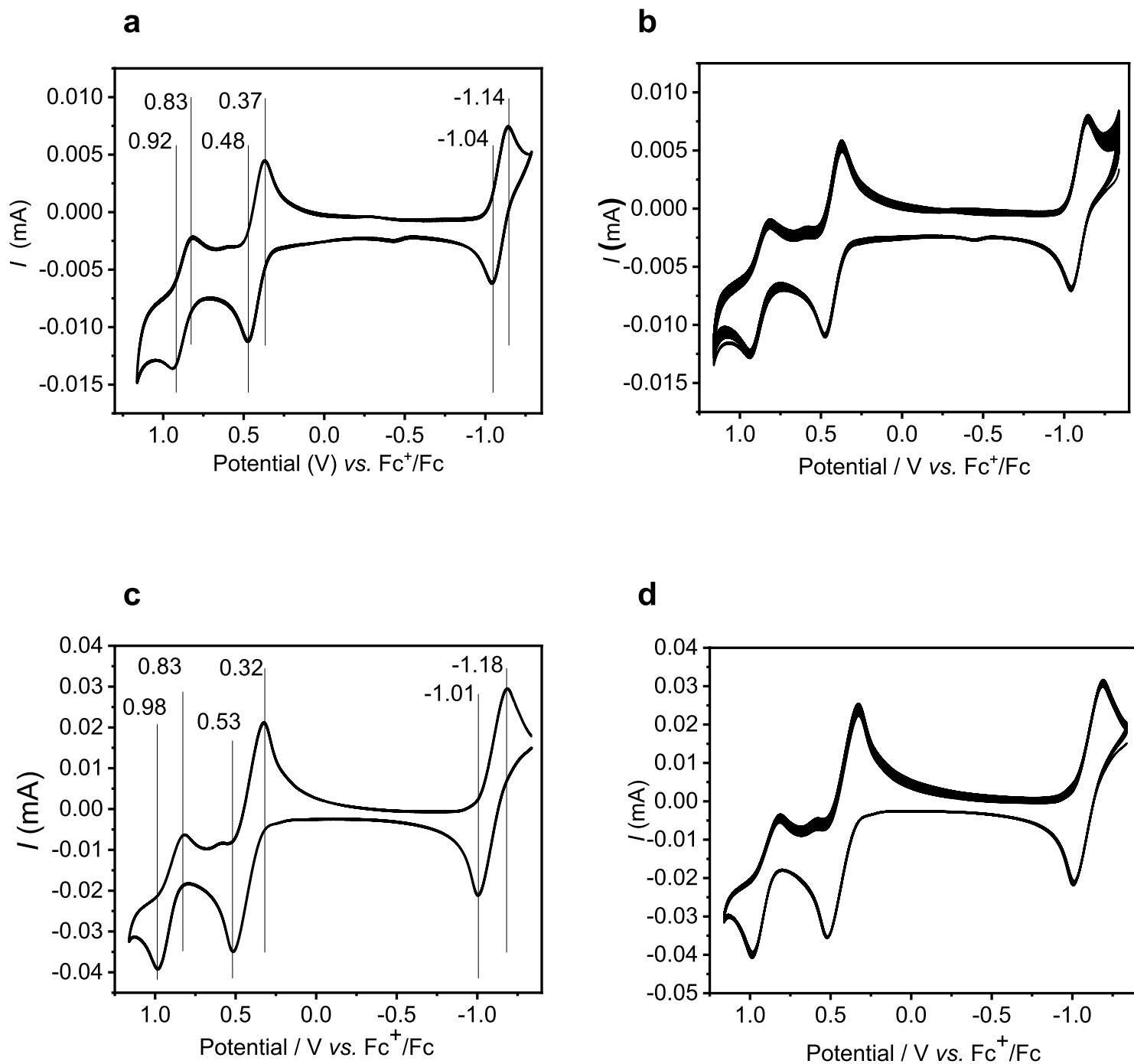
The datasets collected and analysed in this work are available at <https://doi.org/10.17863/CAM.31543>.

a**b**

Extended Data Fig. 1 | Thermal stability and electron paramagnetic resonance measurements of TTM-3NCz and TTM-3PCz.

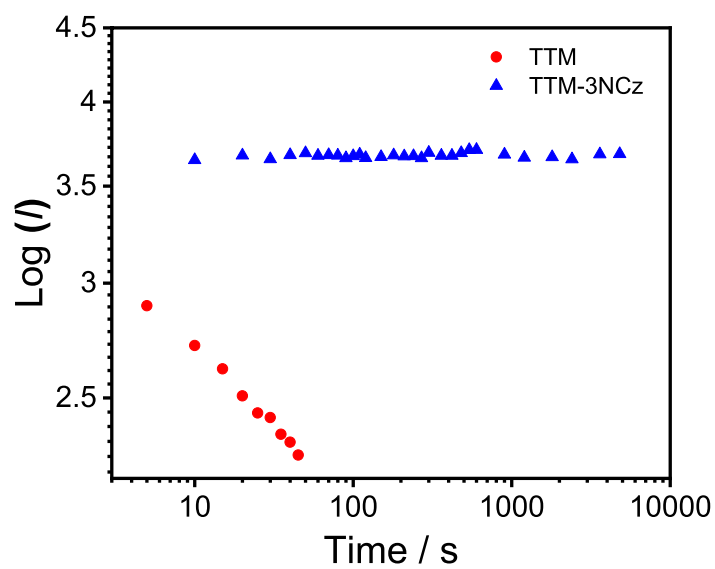
a, Thermogravimetric analysis measurements show thermal

decomposition temperatures of 362 °C (TTM-3NCz) and 367 °C (TTM-3PCz). **b**, EPR spectra for solid samples at room temperature. ESR, electron spin resonance; B , magnetic field.



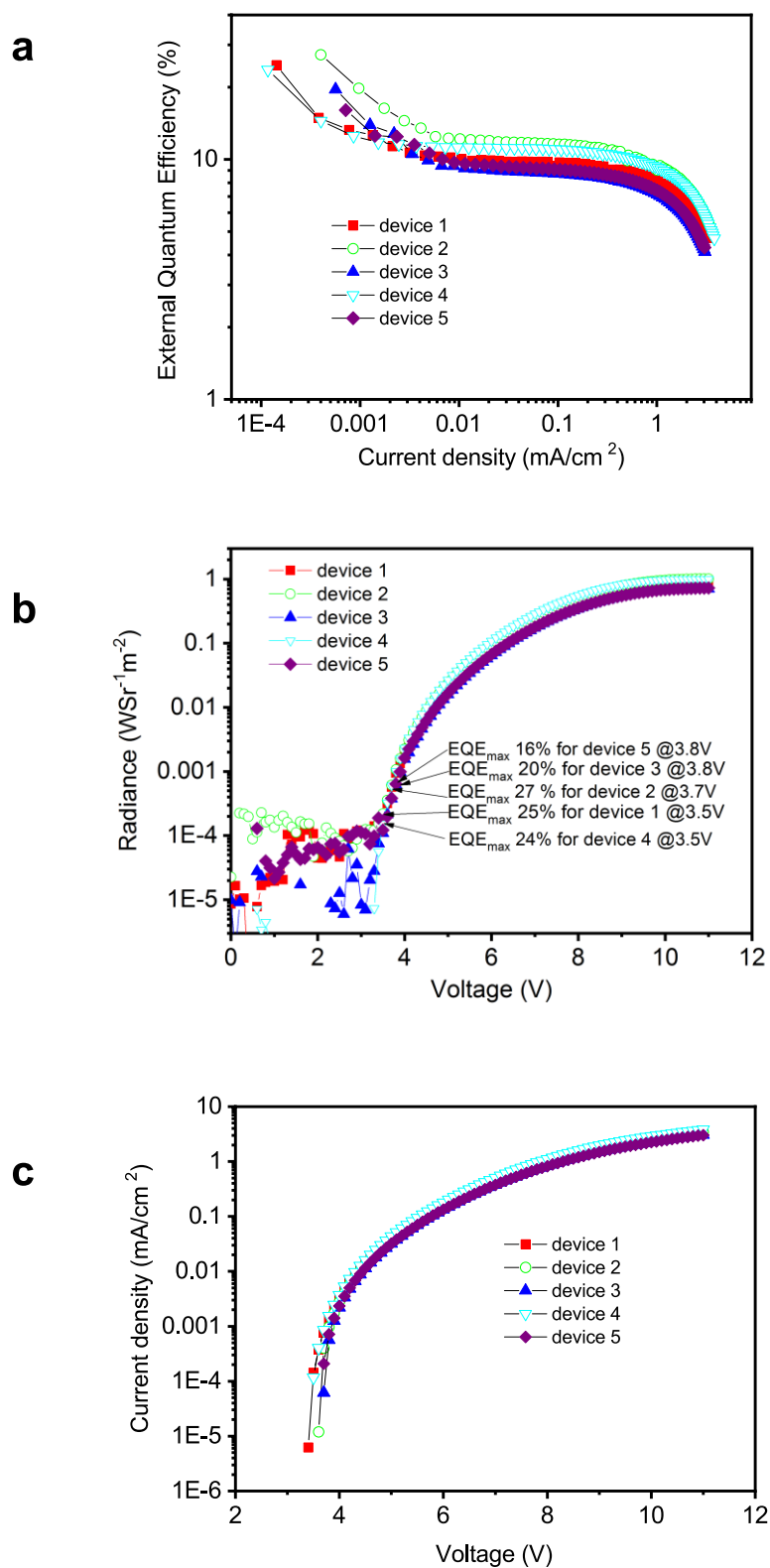
Extended Data Fig. 2 | Electrochemical properties and stability of TTM-3NCz and TTM-3PCz. **a, c,** Cyclic voltammograms of TTM-3NCz (**a**) and TTM-3PCz (**c**) in CH_2Cl_2 . For both TTM-3NCz and TTM-3PCz, the average of the cathodic and anodic potentials gives a reduction potential of -1.1 V, first oxidation potential of $+0.4$ V and

second oxidation potential of $+0.9$ V. **b, d,** Multi-cycle (20 cycles) cyclic voltammetry measurements of TTM-3NCz (**b**) and TTM-3PCz in CH_2Cl_2 (**d**). A ferrocene cation/ferrocene (Fc^+/Fc) reference redox couple was used for the measurements.



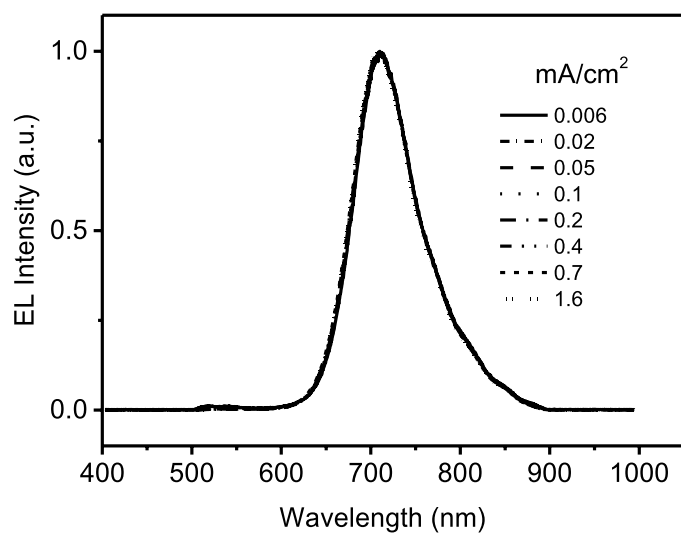
Extended Data Fig. 3 | Photostability of TTM and TTM-3NCz. Luminescence intensity (I) of TTM-3NCz and TTM solutions ($10\text{ }\mu\text{M}$, cyclohexane) as a function of time. A pulsed laser of 355 nm with an

energy density of 315 kW cm^{-2} (pulse width, 8 ns; frequency, 10 Hz) was used under ambient conditions.



Extended Data Fig. 4 | Device reproducibility for TTM-3NCz. **a**, EQE–current density plots for five TTM-3NCz devices. The peak EQE_{max} values are: 25% (device 1), 27% (device 2), 20% (device 3), 24% (device 4) and 16% (device 5). The EQE at 1 mA cm^{-2} is: 8% (device 1), 10% (device 2),

7% (device 3), 9% (device 4) and 7% (device 5). **b**, Radiance–voltage plots for the same five TTM-3NCz devices. The radiance levels for EQE_{max} are indicated, which can be distinguished from the noise. **c**, Current density–voltage plots for the same five TTM-3NCz devices.



Extended Data Fig. 5 | Device stability for TTM-3NCz.

Electroluminescence spectra for TTM-3NCz devices operated at current densities of 0.006–1.6 mA cm^{−2}. There is no current dependence.

Efficient and stable emission of warm-white light from lead-free halide double perovskites

Jiajun Luo^{1,11}, Xiaoming Wang^{2,11}, Shunran Li^{1,11}, Jing Liu^{1,11}, Yueming Guo³, Guangda Niu¹, Li Yao¹, Yuhao Fu⁴, Liang Gao^{1,5}, Qingshun Dong⁶, Chunyi Zhao⁷, Meiying Leng¹, Fusheng Ma⁶, Wenxi Liang¹, Liduo Wang⁶, Shengye Jin⁷, Junbo Han⁸, Lijun Zhang⁴, Joanne Etheridge^{3,9}, Jianbo Wang¹⁰, Yanfa Yan^{2*}, Edward H. Sargent⁵ & Jiang Tang^{1*}

Lighting accounts for one-fifth of global electricity consumption¹. Single materials with efficient and stable white-light emission are ideal for lighting applications, but photon emission covering the entire visible spectrum is difficult to achieve using a single material. Metal halide perovskites have outstanding emission properties^{2,3}; however, the best-performing materials of this type contain lead and have unsatisfactory stability. Here we report a lead-free double perovskite that exhibits efficient and stable white-light emission via self-trapped excitons that originate from the Jahn–Teller distortion of the AgCl_6 octahedron in the excited state. By alloying sodium cations into $\text{Cs}_2\text{AgInCl}_6$, we break the dark transition (the inversion-symmetry-induced parity-forbidden transition) by manipulating the parity of the wavefunction of the self-trapped exciton and reduce the electronic dimensionality of the semiconductor⁴. This leads to an increase in photoluminescence efficiency by three orders of magnitude compared to pure $\text{Cs}_2\text{AgInCl}_6$. The optimally alloyed $\text{Cs}_2(\text{Ag}_{0.60}\text{Na}_{0.40})\text{InCl}_6$ with 0.04 per cent bismuth doping emits warm-white light with 86 ± 5 per cent quantum efficiency and works for over 1,000 hours. We anticipate that these results will stimulate research on single-emitter-based white-light-emitting phosphors and diodes for next-generation lighting and display technologies.

Metal halide perovskites have rapidly advanced the field of optoelectronic devices because of their exceptional defect tolerance, low-cost solution processing and tunable emission across the visible spectrum^{5–8}. For example, the photoluminescence quantum yield (PLQY) of perovskite nanocrystals is now close to unity^{9,10}, and green and red electroluminescent devices have been reported to have external quantum efficiencies that reach 20.1%^{11–14}. For lighting applications, white emission from a single emitter layer is of particular interest, because it simplifies device structure and avoids the self-absorption and colour instability seen in mixed and multiple emitters¹⁵.

Broadband and white emission typically originate from self-trapped excitons (STEs) that exist in semiconductors with localized carriers and a soft lattice^{16–18}. Although hybrid metal halide perovskites, particularly those with low-dimensional crystal structures^{15,19–21}, have received considerable attention as broadband-emission materials, they rarely achieve high PLQY²¹. Further challenges in their use as emitters include their reliance on water-soluble lead-based materials, unsatisfactory stability and a lack of systematic understanding of the origins of white emission.

Here we focused on the double perovskite $\text{Cs}_2\text{AgInCl}_6$, which is a promising material emitting warm-white light, in view of its broad spectrum (400–800 nm) and its all-inorganic and lead-free nature^{22–24}.

We first performed first-principles density-functional-theory and many-body perturbation-theory calculations using the GW approximation and the Bethe–Salpeter equation (BSE) to understand the origins of the broadband emission in $\text{Cs}_2\text{AgInCl}_6$. The GW-BSE calculations indicated that the lowest exciton, which has a binding energy E_b of 0.25 eV, is dark (emits no photons) because the associated transition is parity-forbidden²⁴ (Fig. 1a). This exciton was calculated with the crystal structure fixed in its ground-state equilibrium, which represents the situation of the free exciton. We then investigated exciton–phonon coupling by relaxing the lattice, which represents the situation of the STEs (Fig. 1b). We found that the STEs in $\text{Cs}_2\text{AgInCl}_6$ arise from a strong Jahn–Teller distortion of the AgCl_6 octahedron (see inset of Fig. 1b); that is, the Ag–Cl bonds are elongated by 0.08 Å in the axial direction but compressed by 0.2 Å in the equatorial plane. Hole trapping at Ag atoms that changes the electronic configuration of Ag to $4d^9$ favours a Jahn–Teller distortion. The STE has the same orbital character as the free exciton, indicating a parity-forbidden transition. The self-trapping energy E_{st} and lattice-deformation energy E_d —which are the excited-state and ground-state energy differences between the STE and free-exciton configurations, as shown in the configuration coordinate diagram of Fig. 1c—were calculated to be 0.53 eV and 0.67 eV, respectively. The emission energy was thus calculated to be $E_{PL} = E_g - E_{st} - E_d - E_b = 1.82$ eV, where $E_g = 3.27$ eV is the fundamental band-gap energy, based on GW calculations and experimental results. This value agrees with the experimental photoluminescence peak value²² of 2 eV. The phonon frequency, $\hbar\Omega_g$ (\hbar , reduced Planck constant), of the ground state, obtained by fitting the configuration coordinate diagram, is 18.3 meV, which agrees well with the phonon eigenmode of 17 meV. The corresponding eigenvector shows displacement in agreement with the Jahn–Teller distortion (Extended Data Fig. 1), consistent with the view that the Jahn–Teller distortion is responsible for STE formation in $\text{Cs}_2\text{AgInCl}_6$. Strong electron–phonon coupling, which is necessary for STE formation, is confirmed by the large Huang–Rhys²⁵ factor $S = E_d/\hbar\Omega_g = 37$, consistent with experimental results (Extended Data Fig. 2). With the phonon frequency $\hbar\Omega_e = 17.4$ meV of the excited state, we can estimate the exciton self-trapping time as $\tau = 2\pi/\Omega_e = 238$ fs, which indicates an ultrafast transition from a free exciton to an STE following photoexcitation. The calculated photoluminescence spectrum and a comparison with the experimental data are shown in Fig. 1d. Overall the agreement is good, except for the small deviations at 400–450 nm, which could be attributed to the free-exciton emission not accounted for in our calculations.

The above theoretical analysis indicates an extremely low PLQY for pure $\text{Cs}_2\text{AgInCl}_6$. The PLQY is defined as the ratio of the radiative

¹Sargent Joint Research Center, Wuhan National Laboratory for Optoelectronics (WNLO) and School of Optical and Electronic Information, Huazhong University of Science and Technology (HUST), Wuhan, China. ²Department of Physics and Astronomy and Wright Center for Photovoltaics Innovation and Commercialization, The University of Toledo, Toledo, OH, USA. ³Department of Materials Science and Engineering, Monash University, Clayton, Victoria, Australia. ⁴State Key Laboratory of Superhard Materials, Key Laboratory of Automobile Materials of MOE, and School of Materials Science and Engineering, Jilin University, Changchun, China. ⁵Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada. ⁶Key Laboratory of Organic Optoelectronics and Molecular Engineering of Ministry of Education, Department of Chemistry, Tsinghua University, Beijing, China. ⁷State Key Laboratory of Molecular Reaction Dynamics and Collaborative Innovation Center of Chemistry for Energy Materials (iChEM), Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China. ⁸Wuhan National High Magnetic Field Center, Huazhong University of Science and Technology (HUST), Wuhan, China. ⁹Monash Centre for Electron Microscopy, Monash University, Clayton, Victoria, Australia. ¹⁰School of Physics and Technology, Center for Electron Microscopy, MOE Key Laboratory of Artificial Micro- and Nano-structures, and Institute for Advanced Studies, Wuhan University, Wuhan, China. ¹¹These authors contributed equally: Jiajun Luo, Xiaoming Wang, Shunran Li, Jing Liu. *e-mail: yanfa.yan@utoledo.edu; jtang@mail.hust.edu.cn

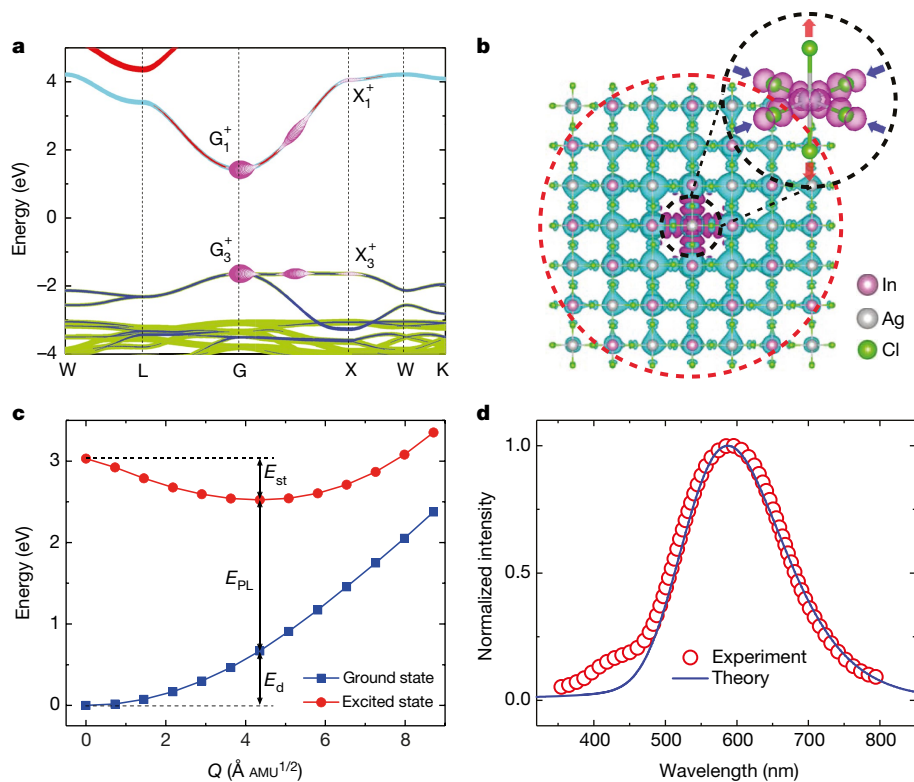


Fig. 1 | Computational studies of the STEs in Cs₂AgInCl₆. **a**, GW band structure of Cs₂AgInCl₆. The orbital characters and free-exciton wavefunction are plotted as a fat-band structure. The green, blue, cyan and red colours denote the Cl 3p, Ag 4d, In 5s and Ag 5s orbitals, respectively. The magenta circles indicate the lowest free-exciton amplitude $|A_{vc\mathbf{k}}|$, where $|S\rangle = \sum_{vc\mathbf{k}} A_{vc\mathbf{k}} |vc\rangle$ is the exciton wavefunction, v and c denote the valence and conduction states, and \mathbf{k} is the wavevector. $|S\rangle$ is derived from the electron and hole states with the same parity (labels at the zone centre G and X) along GX, implying a dark transition. **b**, STE in Cs₂AgInCl₆. Cs atoms are omitted for clarity. The cyan and magenta isosurfaces represent the electron and hole orbital densities ($\rho = \langle \psi | \psi \rangle$), respectively. The

recombination rate (k_{rad}) to the sum of the radiative and non-radiative (k_{non}) recombination rates. From Fermi's golden rule, k_{rad} is proportional to the transition dipole moment, $\mu = \langle \varphi_h | \hat{\mu} | \varphi_e \rangle$, where φ_e and φ_h are the electron and hole wavefunctions, respectively, and $\hat{\mu}$ is the electric dipole operator. The dark transition of the free excitons and STEs in Cs₂AgInCl₆ results in extremely low radiative recombination rates, leading to low PLQY (<0.1%; Extended Data Fig. 2d). Increasing k_{rad} and reducing k_{non} are two strategies to enhance the PLQY. The first and most critical step towards improving the PLQY is to break the parity-forbidden transition by manipulating the symmetry of the STE wavefunction. A practical approach to this end is to partially substitute Ag with an element that can sustain the double-perovskite structure, but has a distinctively different electronic configuration to Ag, such as a group-IA element (alkali metal). We therefore explored alloying Na into Cs₂AgInCl₆. Broadband emission was also observed in pure Cs₂NaInCl₆ (Extended Data Fig. 3), but with very low efficiency due to strong phonon emission, as indicated by a simulated high Huang–Rhys factor of 80 at the excited state. We note that the Huang–Rhys factor can potentially serve as the figure of merit for the design of efficient white-light-emitting materials from STEs (Extended Data Table 1). Because the lattice mismatch between Cs₂NaInCl₆ and Cs₂AgInCl₆ is as low as 0.30% (Supplementary Table 1), we anticipated that Na⁺ could be incorporated uniformly into Cs₂AgInCl₆, without causing detrimental defects or phase separation. For the synthesis, CsCl, NaCl, AgCl and InCl₃ precursors were mixed into an HCl solution in a hydrothermal autoclave, which was heated for a given time and then slowly cooled down, resulting in white precipitates as final products. This straightforward synthesis gave a product yield of nearly 90%.

electron state (red dashed circle) is rather extended and the hole state (black dashed circle) is compact, consistent with the small (large) effective mass of the conduction (valence) band shown in **a**. The inset shows the Jahn–Teller distortion of the AgCl₆ octahedron. Here the hole isosurface is obvious, whereas the electron isosurface is invisible owing to its small density. **c**, Configuration coordinate diagram for the STE formation. E_{st} , E_{d} and E_{PL} are the self-trapping, lattice-deformation and emission energies, respectively. **d**, Calculated photoluminescence spectrum compared with the experimental result. The calculated curve has been shifted to align its maximum with that of the experimentally measured curve for better comparison.

X-ray diffraction (XRD) patterns of a series of compositions (Fig. 2a) confirmed the pure double-perovskite phase. The intensity of the (111) diffraction peak (marked with an asterisk in Fig. 2a) is related to the Na/Ag composition through the dispersion factor of the Na, Ag and In atoms²⁶. These agree well with the compositions determined using inductively coupled plasma optical emission spectrometry (ICP-OES; Supplementary Table 3). This observation also suggests a high degree of B(I) and B'(III) site ordering and negligible antisite defects (Supplementary Fig. 2). The refined lattice parameters follow a linear increase upon Na substitution, indicating solid-solution behaviour with Na⁺/Ag⁺ randomly distributed²⁷ at B(I) sites in Cs₂Ag_xNa_{1-x}InCl₆ (Extended Data Fig. 4). Upon Na alloying, an evident excitonic absorption peak emerged near 365 nm, and the intensity of white emission was enhanced by three orders of magnitude compared to the pure Cs₂AgInCl₆ and Cs₂NaInCl₆ (Fig. 2b). A similar phenomenon was also found in Li-doped Cs₂AgInCl₆ and Na-doped Cs₂AgSbCl₆ (Extended Data Fig. 5), suggesting a general trend of alkali-metal-induced photoluminescence enhancement in double perovskites. We then recorded the photoluminescence spectra of a series of Cs₂Ag_xNa_{1-x}InCl₆ powders by varying the measurement temperatures, and found that the extracted activation energy (Supplementary Figs. 6, 7) increases monotonically with increasing Na content, suggesting suppression of the non-radiative process and thermal quenching upon Na alloying. With optimized Na content, Bi doping and slow cooling, we obtained the highest PLQY of (86 ± 5)% at a Na content of about 40% (Fig. 2c, Supplementary Fig. 8). To the best of our knowledge, this PLQY represents the highest efficiency reported for white-emitting materials (Supplementary Table 4). The best-

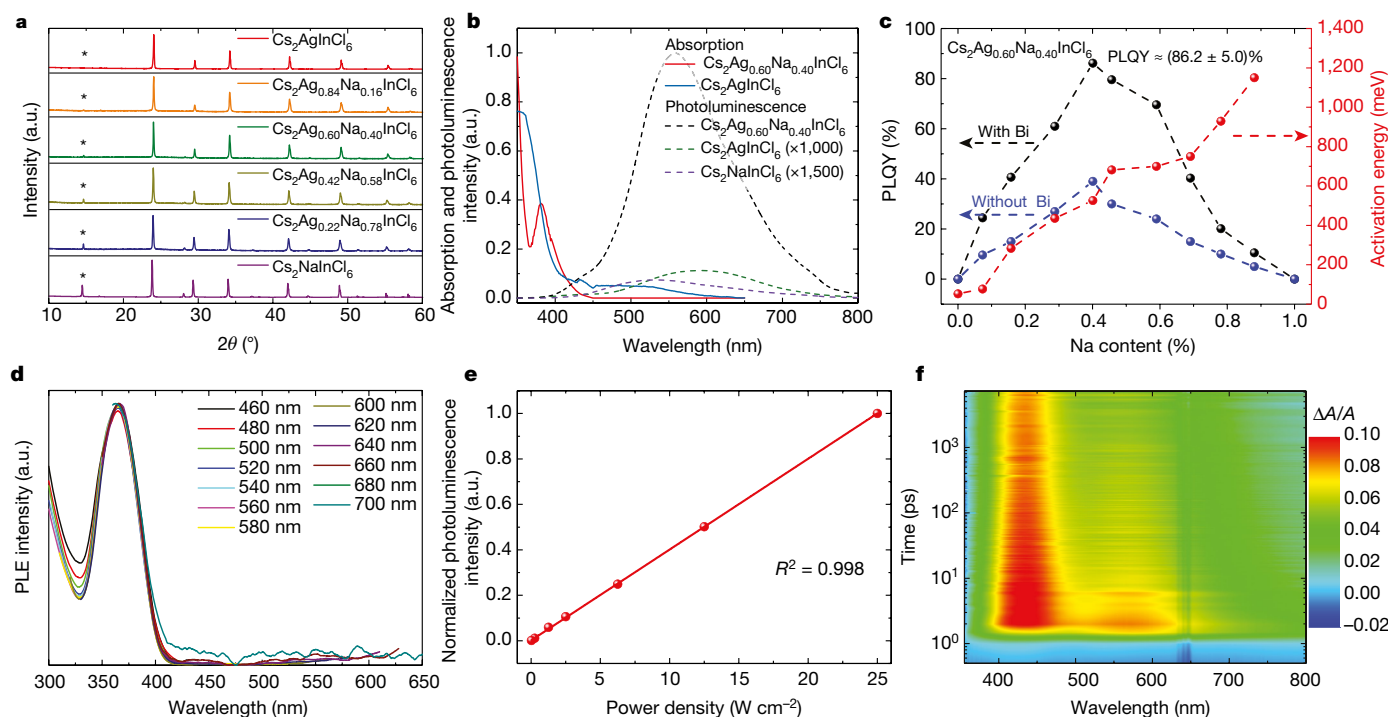


Fig. 2 | Characterization of $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$ with different Na content. All samples were doped using a small amount (0.04%, atomic ratio to In) of Bi, and the compositions were determined from ICP-OES results (Supplementary Table 3). **a**, XRD patterns of $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$ powders with different Na content. The asterisk marks the (111) diffraction peak. **b**, Optical absorption (solid lines) and photoluminescence (dashed lines) spectra of pure $\text{Cs}_2\text{AgInCl}_6$ and $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$. **c**, Activation energy and PLQY of $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$ powder versus Na content. The reproducibility of the PLQY results is shown in Supplementary Fig. 8d (best, about 86.2%);

average, about 71.0%; lowest, about 56.0%). The dashed lines are guides for the eye. **d**, Excitation spectra of photoluminescence, measured at different wavelengths. **e**, Emission intensity versus excitation power for $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$. The linear fit result has a high R^2 value of 0.998. **f**, Transient absorption spectra for $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$ (laser pulse of 325 nm and $4 \mu\text{J cm}^{-2}$). $\Delta A/A$ is the optical density. The irregular peaks located at about 650 nm are from frequency doubling of the pumping light. All data shown in the figure were obtained from measurements on $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$ powder and crystals.

performing white-light-emitting lead halide perovskites $\text{C}_6\text{N}_2\text{H}_{14}\text{PbBr}_4^{28}$ and $\text{CuGaS}_2/\text{ZnS}$ quantum dots²⁹ exhibit PLQYs of 20% and 73%, respectively. The Bi^{3+} incorporation is believed to improve crystal perfection and promote exciton localization³⁰, further enhancing the PLQY (Extended Data Fig. 6).

The STE origin of the white emission was further experimentally confirmed via photoluminescence excitation (PLE) spectra (Fig. 2d). For emission from 460 to 700 nm, the PLE spectra exhibit identical shapes and features, indicating that the white emission originates from the relaxation of the same excited state. The experimental observations that the PLE spectra decrease to nearly zero at wavelengths above 400 nm, that the emission intensity from $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$ exhibits a linear dependence on the excitation power (Fig. 2e) and that the PLQY results are independent from the photoexcitation power (Supplementary Fig. 9) all suggest that the emission does not arise from permanent defects. Surface-defect emission is also ruled out by the comparable photoluminescence intensity of single crystals and ball-milled powders (Supplementary Fig. 10). The transient absorption data further provide direct evidence of STEs¹⁷. With 325-nm-wavelength laser photoexcitation, $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$ exhibited a broad photoinduced absorption at energies across the visible spectrum (Fig. 2f, Supplementary Fig. 11), with an onset time of about 500 fs, consistent with our calculated exciton self-trapping time.

We performed further theoretical analysis to understand the trend of the PLQY as a function of Na content. In Fig. 3a, we show the calculated transition dipole moment of $\text{Cs}_2\text{Ag}_{1-x}\text{Na}_x\text{InCl}_6$ as a function of Na concentration. It is clear that with the increase of Na content, the transition dipole moment first increases and then decreases, reflecting the observed composition-dependent PLQY. Figure 3b compares the electron wavefunction of the STEs before and after the Na alloying. Na incorporation breaks the inversion symmetry of the $\text{Cs}_2\text{AgInCl}_6$ lattice and changes the electron wavefunction at the Ag site from symmetric to

asymmetric; this results in a parity change in the STE wavefunction and consequently allows radiative recombination. Because Na^+ contributes to neither the conduction-band minimum nor the valence-band maximum of the alloy, the second effect of Na incorporation is to reduce the electronic dimensionality⁴ of the $\text{Cs}_2\text{AgInCl}_6$ lattice by partially isolating the AgCl_6 octahedra (Supplementary Fig. 12). The newly formed NaCl_6 octahedra serve as barriers that confine the spatial distribution of the STEs (Fig. 3c), thus enhancing the electron and hole orbital overlap and increasing the transition dipole moment. For example, the radius of the STE is reduced from more than 20 Å for the pure $\text{Cs}_2\text{AgInCl}_6$ to only 9 Å with 50% Na incorporation, which increases the transition dipole moment from zero to 0.07 (in arbitrary units, a.u.).

Two factors account for the decreased PLQY upon further increasing the Na content. For Na-rich compounds, the electron remains strongly confined within a single In octahedron (In 5s and Cl 3p), and the hole is always located on the Ag 4d orbital and the neighbouring Cl 3p orbitals (Fig. 3d). Therefore, the orbital spatial overlap between electrons and holes for the STEs, and hence the transition dipole moment, is markedly reduced. The second factor is the increased non-radiative loss in the Na-rich alloy. We found that the excited- and ground-state curves cross in the configuration coordinate diagram of pure $\text{Cs}_2\text{NaInCl}_6$ (Fig. 3e), which means that some photoexcited electrons can recombine with holes non-radiatively through phonon emission. The resulting diminished transition dipole moment and enhanced non-radiative recombination rates explain the decreased PLQY for Na-rich alloys.

The photoluminescence spectrum of the best-performing $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$ powder exhibits extended overlap with the sensitivity of the human eye to optical wavelengths (that is, the luminosity function) (Fig. 3a), which enables a theoretical luminous efficacy reaching about 373 lm W^{-1} . Emission stability is another

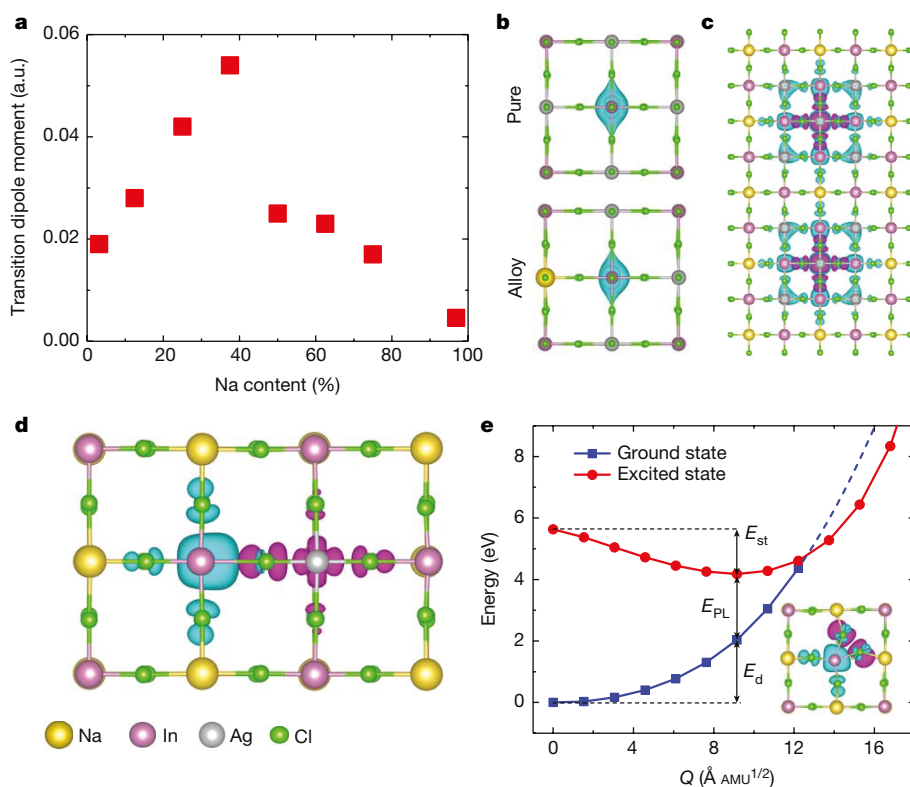


Fig. 3 | Mechanistic investigations of PLQY in Cs₂Ag_{1-x}Na_xInCl₆. **a**, Transition dipole moment, μ , as a function of Na content in Cs₂Ag_{1-x}Na_xInCl₆. Assuming constant non-radiative recombination, the PLQY is proportional to μ . **b**, Parity change of the electron wavefunction (isosurface at Ag site; see the key in the inset of **d**) of the STE before and after Na incorporation. **c**, Configuration showing the strengthened STE confinement by the surrounding NaCl₆ octahedra. The STEs are confined within two lattice parameters surrounded by the NaCl₆ octahedra. **d**, STE in Na-rich Cs₂Ag_{1-x}Na_xInCl₆. The STE is located in two neighbouring octahedra (AgCl₆ and InCl₆) with the hole derived from the Ag 4d/Cl 3p orbitals and the electron from In 5s/Cl 3p orbitals. **e**, Configuration coordinate diagram of the STE formation in Cs₂NaInCl₆ (inset). The STE is located within a single distorted InCl₆ octahedron. The hole is located at the well known V_k centre, that is, a Cl₂⁻ dimer ion, whereas the electron is derived from In 5s/Cl 3p orbitals. The separation of the electron and hole makes the optical transition very weak. In **b–e**, the cyan and magenta isosurfaces denote electrons and holes, respectively.

key, yet very challenging, parameter for lighting applications. The Cs₂Ag_{0.60}Na_{0.40}InCl₆ materials demonstrated little emission degradation when tested from 233 K to 343 K. A version of the material slightly richer in Na (Na/(Ag + Na) = 0.46) showed stable emission up to 393 K (Supplementary Fig. 13). We further annealed our Cs₂Ag_{0.60}Na_{0.40}InCl₆ powders on a hotplate at 150 °C for 1,000 h and observed little photoluminescence decay of the white emission (Fig. 4b). We propose that the strongly bound excitons and nearly defect-free lattice of Cs₂Ag_{0.60}Na_{0.40}InCl₆ prevent photoluminescence quenching and that

the all-inorganic composition also helps resist thermal stress (decomposition temperature of up to about 863 K; Supplementary Fig. 14).

We fabricated a white-emission light-emitting diode (LED) by directly pressing the Cs₂Ag_{0.60}Na_{0.40}InCl₆ powders onto a commercial ultraviolet LED chip, without using epoxy or silica encapsulation for protection. With the contribution from the blue light of the ultraviolet LED chip (380–410 nm), the device has CIE coordinates (0.396, 0.448), located at a warm-white point with a correlated colour temperature of 4,054 K, which fulfils the requirements for indoor lighting. The

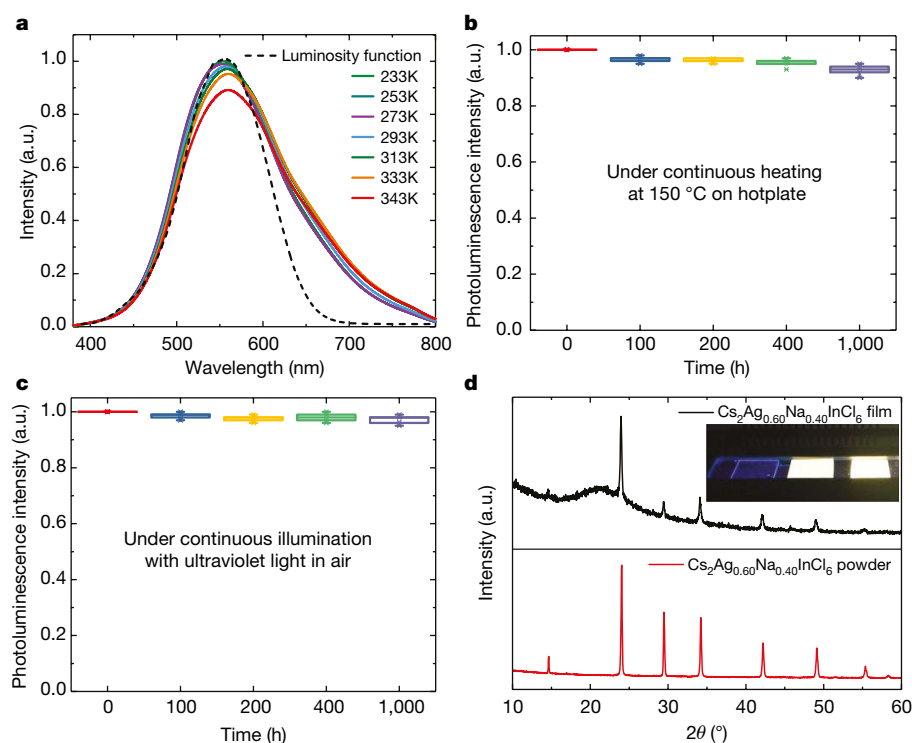


Fig. 4 | White emission from Cs₂Ag_{1-x}Na_xInCl₆. **a**, Luminosity function (dashed line) and photoluminescence spectra (solid lines) of Cs₂Ag_{0.60}Na_{0.40}InCl₆ measured at different temperatures from 233 K to 343 K. **b**, Photoluminescence stability of Cs₂Ag_{0.60}Na_{0.40}InCl₆ against continuous heating at 150 °C on a hotplate, measured after cooling to room temperature. **c**, Operational stability of Cs₂Ag_{0.60}Na_{0.40}InCl₆ down-conversion devices, measured in air without any encapsulation. The box plot shows the results for five different samples measured separately, with the box edges representing quartiles, the band inside the box showing the median and the end of the whiskers representing the minimum and maximum of the data. **d**, XRD patterns of a Cs₂Ag_{0.60}Na_{0.40}InCl₆ film (black line) and powder (red line). The inset shows a 300-nm-thick quartz substrate and 500-nm-thick Cs₂Ag_{0.60}Na_{0.40}InCl₆ films under 254-nm ultraviolet illumination.

white LED showed negligible degradation when operated at about 5,000 cd m⁻² for over 1,000 h in air (Fig. 4c). This outstanding photometric performance, combined with its easy manufacture, indicate promise for white-phosphor applications.

The broadband emission associated with the STEs provides a new strategy to produce single-material-based, white-light electroluminescence. We thus fabricated prototype double-perovskite-based electroluminescence devices. XRD measurements confirmed the pure phase of the thermally evaporated Cs₂Ag_{0.60}Na_{0.40}InCl₆ film, which showed bright and uniform warm-white photoluminescence under ultraviolet-lamp excitation (Fig. 4d). Our electroluminescence device demonstrated bias-insensitive broadband emission and a peak current efficiency of 0.11 cd A⁻¹, which was mainly limited by the low quality of the Cs₂Ag_{0.60}Na_{0.40}InCl₆ films (Supplementary Figs. 15–17). Further research should focus on optimizing emitting-layer quality and device configuration to increase electroluminescence performance.

In summary, Na alloying into Cs₂AgInCl₆ breaks the parity-forbidden transition and reduces its electronic dimensionality, leading to efficient white emission via radiative recombination of STEs. This white-light-emitting material also demonstrates outstanding stability and low-cost manufacture, indicating promise for solid-state lighting. We believe that halide double perovskites hold great potential for display and lighting applications and merit further study to realize their full potential.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <http://sci-hub.tw/10.1038/s41586-018-0691-0>.

Received: 27 February 2018; Accepted: 31 August 2018;

Published online 7 November 2018.

- Sun, Y. et al. Management of singlet and triplet excitons for efficient white organic light-emitting devices. *Nature* **440**, 908–912 (2006).
- Tan, Z. K. et al. Bright light-emitting diodes based on organometal halide perovskite. *Nat. Nanotechnol.* **9**, 687–692 (2014).
- Cho, H. et al. Overcoming the electroluminescence efficiency limitations of perovskite light-emitting diodes. *Science* **350**, 1222–1225 (2015).
- Xiao, Z. et al. Searching for promising new perovskite-based photovoltaic absorbers: the importance of electronic dimensionality. *Mater. Horiz.* **4**, 206–216 (2017).
- Kojima, A., Teshima, K., Shirai, Y. & Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.* **131**, 6050–6051 (2009).
- Burschka, J. et al. Sequential deposition as a route to high-performance perovskite-sensitized solar cells. *Nature* **499**, 316–319 (2013).
- Lee, M. M., Teuscher, J., Miyasaka, T., Murakami, T. N. & Snaith, H. J. Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. *Science* **338**, 643–647 (2012).
- Yin, W. J., Shi, T. & Yan, Y. Unique properties of halide perovskites as possible origins of the superior solar cell performance. *Adv. Mater.* **26**, 4653–4658 (2014).
- Protesescu, L. et al. Nanocrystals of cesium lead halide perovskites (CsPbX₃, X=Cl, Br, and I): novel optoelectronic materials showing bright emission with wide color gamut. *Nano Lett.* **15**, 3692–3696 (2015).
- Zhou, Q. et al. In situ fabrication of halide perovskite nanocrystal embedded polymer composite films with enhanced photoluminescence for display backlights. *Adv. Mater.* **28**, 9163–9168 (2016).
- Wang, N. et al. Perovskite light-emitting diodes based on solution-processed self-organized multiple quantum wells. *Nat. Photon.* **10**, 699–704 (2016).
- Yuan, M. et al. Perovskite energy funnels for efficient light-emitting diodes. *Nat. Nanotechnol.* **11**, 872–877 (2016).
- Yang, X. et al. Efficient green light-emitting diodes based on quasi-two-dimensional composition and phase engineered perovskite with surface passivation. *Nat. Commun.* **9**, 570 (2018); correction **9**, 1169 (2018).
- Zhao, B. et al. High-efficiency perovskite-polymer bulk heterostructure light-emitting diodes. Preprint at <https://arxiv.org/abs/1804.09785> (2018).
- Dohner, R. E., Hoke, T. K. & Karunadasa, I. H. Self-assembly of broadband white-light emitters. *J. Am. Chem. Soc.* **136**, 1718–1721 (2014).
- Song, K. S. & Williams, R. T. *Self-Trapped Excitons* (Springer, New York, 2008).
- Smith, M. D. & Karunadasa, I. H. White-light emission from layered halide perovskites. *Acc. Chem. Res.* **51**, 619–627 (2018).
- Ueta, M., Kanzaki, H., Kobayashi, K., Toyozawa, Y. & Hanamura, E. in *Excitonic Processes in Solids* 309–369 (Springer, Berlin, Heidelberg, 1986).
- Dohner, R. E., Jaffe, A., Bradshaw, R. L. & Karunadasa, I. H. Intrinsic white-light emission from layered hybrid perovskites. *J. Am. Chem. Soc.* **136**, 13154–13157 (2014).
- Mao, L., Wu, Y., Stoumpos, C. C., Wasielewski, M. R. & Kanatzidis, M. G. White-light emission and structural distortion in new corrugated two-dimensional lead bromide perovskites. *J. Am. Chem. Soc.* **139**, 5210–5215 (2017).
- Zhou, C. et al. Luminescent zero-dimensional organic metal halide hybrids with near-unity quantum efficiency. *Chem. Sci.* **9**, 586–593 (2018).
- Volonakis, G. et al. Cs₂InAgCl₆: a new lead-free halide double perovskite with direct band gap. *J. Phys. Chem. Lett.* **8**, 772–778 (2017).
- Zhao, X. G. et al. Cu–In halide perovskite solar absorbers. *J. Am. Chem. Soc.* **139**, 6718–6725 (2017).
- Meng, W. et al. Parity-forbidden transitions and their impact on the optical absorption properties of lead-free metal halide perovskites and double perovskites. *J. Phys. Chem. Lett.* **8**, 2999–3007 (2017).
- Huang, K. & Rhyas, A. Theory of light absorption and non-radiative transitions in F-centres. *Proc. R. Soc. Lond. A* **204**, 406–423 (1950).
- Lim, T.-W. et al. Insights into cationic ordering in Re-based double perovskite oxides. *Sci. Rep.* **6**, 19746 (2016).
- Maughan, A. E. et al. Defect tolerance to intolerance in the vacancy-ordered double perovskite semiconductors Cs₂SnI₆ and Cs₂TeI₆. *J. Am. Chem. Soc.* **138**, 8453–8464 (2016).
- Yuan, Z. et al. One-dimensional organic lead halide perovskites with efficient bluish white-light emission. *Nat. Commun.* **8**, 14051 (2017).
- Kim, J.-H. et al. White electroluminescent lighting device based on a single quantum dot emitter. *Adv. Mater.* **28**, 5093–5098 (2016).
- Moser, F. & Lyu, S. Luminescence in pure and I-doped AgBr crystals. *J. Lumin.* **3**, 447–458 (1971).

Acknowledgements This work was financially supported by the National Natural Science Foundation of China (51761145048 and 61725401), the National Key R&D Program of China (2016YFB0700702, 2016YFA0204000 and 2016YFB0201204), the HUST Key Innovation Team for Interdisciplinary Promotion (2016JCTD111) and the Program for JLU Science and Technology Innovative Research Team. The calculation of broadband emission at the University of Toledo was supported by the Center for Hybrid Organic Inorganic Semiconductors for Energy (CHOISE), an Energy Frontier Research Center funded by the Office of Basic Energy Sciences, Office of Science within the US Department of Energy. The analysis of the electronic properties of halide double perovskites was funded by the Office of Energy Efficiency and Renewable Energy (EERE), US Department of Energy, under award number DE-EE0006712. Part of the code development was supported by the National Science Foundation under contract number DMR-1807818. Y.Y. acknowledges support from the Ohio Research Scholar Program. For the theoretical calculations we used the resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under contract number DE-AC02-05CH11231. Y.G. and J.E. acknowledge financial support by the Australian Research Council (DP150104483) and the use of instrumentation at the Monash Centre for Electron Microscopy. The authors from HUST thank the Analytical and Testing Center of HUST and the facility support of the Center for Nanoscale Characterization and Devices, WNLO. We also thank Z. Xiao for useful discussion about emission mechanisms and some XRD measurements, as well as T. Zhai, H. Song, Y. Zhou, H. Han, X. Lu and L. Xu for providing access to some facilities.

Reviewer information Nature thanks C. C. Stoumpos and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.T. conceived the idea and guided the whole project. J. Luo, S.L. and J. Liu designed and performed most of the experiments and analysed the data; X.W. performed most of the theoretical calculations and analysis (GW-BSE, STE, photoluminescence) under the guidance of Y.Y.; S.L. discovered the phosphor; L.Y. contributed in electroluminescence device optimization; L.G. carried out transient-absorption experiments; M.L. assisted in data analysis and photoluminescence measurements; Y.G. and J.E. carried out the electron microscopy measurements and analysed the results; Y.F. and L.Z. simulated the band alignment and the contour plots of the valence-band maximum and conduction-band maximum charge densities; C.Z. and S.J. provided some optical measurements; Q.D., F.M., L.W., W.L. and J.H. helped in the PLQY measurement and electroluminescence device fabrication; G.N. was involved in data analysis and experimental design; J.W. contributed to DFT calculations, Y.Y. helped in manuscript writing; J. Luo, X.W., E.H.S. and J.T. wrote the paper; all authors commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0691-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0691-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to Y.Y. or J.T.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Materials. Caesium chloride (CsCl, 99.99%), silver chloride (AgCl, 99.99%), sodium chloride (NaCl, 99.99%), lithium chloride (LiCl, 99.9%), anhydrous indium chloride (InCl₃, 99.999%), anhydrous bismuth chloride (BiCl₃, 99.999%), anhydrous antimony chloride (SbCl₃, 99.99%), zinc acetate dehydrate (Zn(CH₃COO)₂·2H₂O >98%), tetramethylammonium hydroxide (TMAH, 98%) and polyethylenimine (PEIE) were purchased from Sigma Aldrich. Molybdenum oxide (MoO₃, 99%) and 4,4'-cyclohexylidenebis[N,N-bis(4-methylphenyl)benzamine] (TAPC, 99%) were purchased from Guangdong Aglaia Optoelectronic Materials Company. Hydrochloric acid, ethanol, acetone, isopropanol, ethyl acetate, 1-butanol and dimethyl sulfoxide (DMSO, 99%) were purchased from Sinopharm Chemical Reagent Company. Patterned indium tin oxide (ITO) glass substrates (sheet resistance, 15 Ω sq⁻¹) were purchased from Guangdong Xiangcheng Technology Company. All materials were used as received.

Synthesis of alloyed double-perovskite materials. Because double perovskites are generally impurity-sensitive, a Teflon autoclave was soaked overnight with aqua regia and high-purity raw materials were used. Then, 1 mmol anhydrous InCl₃, 0.005 mmol anhydrous BiCl₃ and 2 mmol CsCl were first dissolved in 10 ml of a 10 M HCl solution in a 25-ml Teflon autoclave. Then x mmol of AgCl and $1-x$ mmol of NaCl were added and the solution was heated at 180 °C for 12 h in a stainless-steel Parr autoclave. The solution was then steadily cooled to 50 °C at a speed of 3 °C h⁻¹ (the cooling process was key in determining the PLQY of the products). The as-prepared crystals were then filtered out, washed with isopropanol and dried in a furnace at 60 °C. Na-doped Cs₂AgSbCl₆ was synthesized by substituting the InCl₃ with SbCl₃, and Li-doped Cs₂AgInCl₆ was obtained by mixing 20 mmol LiCl with 1 mmol Cs₂AgInCl₆ in a 25-ml Teflon autoclave containing 4 ml of a 10 M HCl solution, and then following exactly the same procedure as for the Cs₂Ag _{x} Na _{$1-x$} InCl₆ synthesis.

Characterization and calibration of the PLQY of Cs₂Ag _{x} Na _{$1-x$} InCl₆. The PLQY was measured using an absolute photoluminescence measurement system (Hamamatsu Quantaurus-QY) at Tsinghua University. The excitation wavelength was 365 nm, and the step increments and integration time were 1 nm and 0.5 s per data point, respectively. Commercial YAG:Ce³⁺ powder purchased from Hunan LED Company with a standard PLQY (80%–85%, 460 nm excitation) was used to calibrate the system.

Electroluminescence device fabrication. Colloidal ZnO nanocrystals were synthesized following a published procedure³¹. Patterned indium-doped ITO substrates were cleaned by sequential sonication in acetone, ethanol and deionized water, for 30 min in each bath. After drying, solutions of PEIE in isopropanol (0.1 wt%) were spin-coated onto the ITO substrates at 5,000 r.p.m. for 60 s, followed by a layer of ZnO nanocrystals spun at 3,000 r.p.m. for 60 s, and a further PEIE layer (0.1 wt% in 2-methoxyethanol), spin-coated at 5,000 r.p.m. for 60 s. The Cs₂Ag_{0.60}Na_{0.40}InCl₆ film was deposited by thermal evaporation of CsCl, AgCl, NaCl, InCl₃ and BiCl₃ in separate crucibles at a stoichiometric molar ratio of 2:0.6:0.4:1:~0.005. The evaporation rate was monitored by a quartz microbalance. After the pressure of the evaporator chamber (Fangsheng Technology, OMV-FS300) was pumped down to 6×10^{-6} mTorr, one precursor was heated slowly to achieve a desirable deposition rate (CsCl, 0.10–0.20 Å s⁻¹; NaCl, 0.01–0.03 Å s⁻¹; AgCl, 0.05–0.10 Å s⁻¹; InCl₃, 0.10–0.20 Å s⁻¹; BiCl₃, 0.01 Å s⁻¹). The shutter was then manually opened until a certain thickness was deposited. The evaporation sequence was CsCl, InCl₃, BiCl₃, NaCl and AgCl. Then, the Cs₂Ag_{0.60}Na_{0.40}InCl₆ film was exposed to air for 5 min and further annealed at 150 °C in N₂ for 5 min to promote crystallization. Afterwards, 40-nm-thick TAPC layers were deposited at a speed of 0.10–0.20 Å s⁻¹, followed by the deposition of the MoO₃/Al electrode to complete the device (device area, 4 mm²).

Material characterization. Powder XRD measurements were performed by grinding Cs₂Ag _{x} Na _{$1-x$} InCl₆ crystals into fine powders in a mortar, using a Philips X'pert pro MRD diffractometer with Cu K α radiation. High-resolution XRD measurements were conducted on a powder diffractometer (D8 ADVANCE, Bruker) using a Cu K α rotating anode. The absorption and reflectance spectra were measured on an ultraviolet–visible spectrophotometer (PerkinElmer Instruments, Lambda 950) with an integrating sphere, which was calibrated by measuring a reference material (MgO powder) at the same time. The photoluminescence and PLE measurements were carried out using an Edinburgh Instruments Ltd UC920 spectrometer. The temperature-dependent photoluminescence spectra were measured using a Horiba Jobin Yvon LabRAM HR800 Raman spectrometer excited by a 325-nm-wavelength He–Cd laser and at a temperature ranging from 80 to 500 K, achieved using a liquid-nitrogen cooler. The intensity-dependent photoluminescence measurement was also carried out using a picosecond-pulse diode laser (Light Conversion, Pharos) with 365-nm output wavelength and 50-ps pulse width, and the pulse intensity was monitored by a power meter (Ophir PE10BF-C). The power density was controlled by neutral-density filters (Light Conversion, Pharos). The photoluminescence lifetime measurement was performed using time-correlated single-photon counter technology. The excitation beam was a picosecond-pulse diode

laser (Light Conversion, Pharos) with 365-nm output wavelength and 50-ps pulse width. For the transient-absorption measurement, an amplified Yb:KGW laser (Light Conversion, Pharos) with 5-kHz repetition rate was used to generate femtosecond-laser pulses (pump wavelength, 325 nm; intensity, 4 μ J cm⁻²). A crystal with a size of about 0.2 \times 1.0 \times 1.0 mm³ was placed on the glass substrate during the measurement. ICP-OES measurements were carried out using a Perkin Elmer Optima 7300DV spectrometer with the Cs₂Ag _{x} Na _{$1-x$} InCl₆ powders dissolved in HCl. Thermal gravimetric analysis was performed with a PerkinElmer Diamond TG/DTA6300 system at a heating rate of 10 °C min⁻¹ from room temperature to 800 °C in N₂ flow using an alumina crucible. A Cs₂Ag_{0.60}Na_{0.40}InCl₆ thin film fabricated by thermal evaporation was characterized by scanning electron microscopy (FEI Nova NanoSEM450, without Pt coating), ultraviolet photoemission spectroscopy (Specs UVLS, He I excitation, 21.2 eV; referenced to the Fermi edge of argon-etched gold). Stability against heat was measured by simply putting the powders on a 150 °C hotplate in N₂, and the photoluminescence intensity was measured after a certain time interval. We note that all measurements were performed on powder and crystals, except for the electroluminescence measurements, which were made on films.

Transmission electron microscopy analysis. Transmission electron microscopy (TEM) specimens were prepared by crushing the as-grown single crystals and then drop-casting them onto a TEM copper grid covered by an ultrathin carbon film. TEM characterization was carried out on a JEOL 2100F TEM with a field-emission gun operating at 200 kV at the Monash Centre for Electron Microscopy (MCEM). Low-dose selected-area electron diffraction and scanning electron nanobeam diffraction were performed to avoid beam damage. Using a nominal current density of 2 pA cm⁻², no change in lattice parameters was observed after several minutes' exposure. For the scanning electron nanobeam diffraction measurement, a step size of 5 nm was used and a dataset of 10 \times 10 diffraction patterns of (2,048 pixels) \times (2,048 pixels) was collected from a square region of 50 \times 50 nm². We deployed a digital micrograph script for automatic control of the scanning coils and pattern acquisition, developed by J. M. Zuo's group at the University of Illinois at Urbana-Champaign.

First-principles density functional theory, many-body perturbation theory and BSE calculations. Density functional theory, GW and BSE calculations were performed using the VASP code^{32,33} with projector augmented-wave (PAW)³⁴ potentials. A kinetic energy cutoff of 520 eV and Γ -centred $4 \times 4 \times 4$ k -mesh were employed. Because the band gaps of both Cs₂AgInCl₆ and Cs₂NaInCl₆ were found to be sensitive to the bond length, we used the more accurate PBE0³⁵ functional to relax the atomic coordinates with a force tolerance of 0.01 eV Å⁻¹ while keeping the lattice parameters fixed at their experimental values. With the relaxed coordinates, GW calculations were performed using the PBE³⁶ wavefunction. Partial self-consistency on Green's function only—the GW₀ scheme—was adopted. For the GW calculations, an energy cutoff of 200 eV for the response function, 200 real frequency grids for the dielectric function and 1,000 bands were used. The results were further extrapolated to infinite-basis sets and a number of bands³⁷. GW band structures were obtained using Wannier interpolation with the wannier90 code³⁸. BSE calculations were performed using the GW quasiparticle energies. The number of occupied/virtual states used for Cs₂AgInCl₆ and Cs₂NaInCl₆ were 2/2 and 24/4, respectively, to achieve convergence of the several low-lying exciton states. The exciton binding energies were extrapolated to infinitely dense k -meshes. For a finer k -mesh, GW calculations are computationally prohibitive. We used Wannier interpolation to interpolate the GW quasiparticle energies and model the dielectric function³⁹ ϵ_q , which was fitted from the value obtained with a coarser grid to interpolate the dielectric function ($\epsilon_q = 1 + [(\epsilon_\infty - 1)^{-1} + aq^2 + bq^4]^{-1}$, where ϵ_∞ is the static dielectric constant, q is the wave vector, and a and b are fitting parameters).

STE calculation. To study the STE properties, we used the restricted open-shell Kohn–Sham (ROKS) theory^{40–42}, as implemented in the *cp2k* code⁴³. A supercell with a single Γ point was used in the calculation. The double-zeta valence polarization molecularly optimized basis sets⁴⁴, PBE exchange–correlation functional and Goedecker–Teter–Hutter pseudopotentials⁴⁵ were used. Energy cutoffs of 300 Ry and 1,200 Ry (1 rydberg, 1 Ry = 13.605 eV) were used for Cs₂AgInCl₆ and Cs₂NaInCl₆, respectively. The delocalization error of the PBE functional was removed using the scaled Perdew–Zunger self-interaction correction^{46,47} only on the unpaired electrons⁴⁸. The scaling parameter α of the Hartree energy was fitted to reproduce the exciton binding energies calculated by the GW-BSE approach. The exciton binding energy within the ROKS framework is calculated as $E_b = E_g - (S_1 - E_0)$, where S_1 and E_0 are the first excited singlet-state and ground-state energies, respectively. We obtained $\alpha = 0.30$ and $\alpha = 0.34$ for Cs₂AgInCl₆ and Cs₂NaInCl₆, respectively. Because the present self-interaction correction scheme is not meant to correct the bandgap, the excited-state curves in the configuration coordinate diagrams were shifted by aligning the free-exciton energy with that from the GW-BSE calculations. A supercell with a size of 21.0 \times 21.0 \times 21.0 Å³ was found enough to obtain convergence of both the free exciton and the STE of

Cs₂NaInCl₆. However, for Cs₂AgInCl₆, owing to the small effective mass of the electron, a supercell with a size as large as $41.9 \times 41.9 \times 41.9 \text{ \AA}^3$ is needed. For a completely delocalized state, the self-interaction correction is zero; hence, we neglected the self-interaction correction on the electron wavefunction of the STE in Cs₂AgInCl₆. This can safely reduce the supercell size to only $20.9 \times 20.9 \times 20.9 \text{ \AA}^3$. For the alloyed double perovskites, we used a supercell of $21.0 \times 21.0 \times 21.0 \text{ \AA}^3$ and $\alpha = 0.34$.

Configuration coordinate diagram and photoluminescence spectra calculation. The configuration coordinate (Q) diagram was constructed by linearly interpolating the coordinates between the free-exciton and STE configurations and then calculating both the ground-state and excited-state energies at each coordinate. The coordinate difference between the free-exciton and STE configurations is $\Delta Q = \sqrt{\sum_{\kappa,i} M_{\kappa} (R_{\kappa,i}^e - R_{\kappa,i}^g)^2}$, where κ denotes the atom, $i = (x, y, z)$, M is the atomic mass and R are the atomic coordinates with e and g for the excited and ground state, respectively. The calculated ΔQ is $4.35 \text{ \AA AMU}^{1/2}$ and $9.16 \text{ \AA AMU}^{1/2}$, respectively, for Cs₂AgInCl₆ and Cs₂NaInCl₆. The coordinate Q was linearly interpolated between 0 and ΔQ . The phonon frequency Ω was obtained by a third-order polynomial fit $((1/2)\Omega^2 Q^2 + \lambda Q^3$; λ is a fitting parameter) of the excited- or ground-state curve. The normalized photoluminescence intensity in the leading order can be written as⁴⁹ $I(h\nu) = C\nu^x A(h\nu)$ ($x = 3$ for dipole-allowed transition, $x = 5$ for dipole-forbidden transition), where $h\nu$ is the photon energy and C is the normalization factor, which includes the transition dipole moments for dipole-allowed transitions, or the magnetic dipole moments and electric quadrupole moments for dipole-forbidden transitions. A is the normalized spectral function, under the Franck–Condon approximation:

$$A(h\nu) = \sum_{m,n} w_m(T) |\langle \chi_{gn} | \chi_{em} \rangle|^2 \delta(E_{ZPL} + \hbar\omega_m - \hbar\omega_n - h\nu) \quad (1)$$

$w_m(T)$ is the thermal occupation factor of the excited-state phonons with energy $\hbar\omega_m = (m)\hbar\Omega_m$, where Ω_m is the phonon frequency, m is the corresponding quantum number, n denotes the related ground-state quantity, T is the temperature and k_B is the Boltzmann constant. E_{ZPL} is the zero-phonon line energy, which is the energy difference between the minima of the excited- and ground-state curves plus the zero-point energy difference, $(1/2)\hbar(\Omega_e - \Omega_g)$. χ_{em} and χ_{gn} are the harmonic phonon wavefunctions of the excited and ground states, respectively. The Franck–Condon factors $|\langle \chi_{gn} | \chi_{em} \rangle|^2$ were calculated by the recurrence method⁵⁰. The δ function in equation (1) was replaced by the Lorentzian with a broadening parameter of 0.03 eV, which is around the phonon cutoff frequency of Cs₂AgInCl₆ (Extended Data Fig. 2).

LED devices on ultraviolet chips. GaN-based ultraviolet chips (14 W output, 365–370 nm peak emission) were purchased from Taiwan Epileds Company. The Cs₂Ag_{0.60}Na_{0.40}InCl₆ crystals were ball-milled into fine powder, and the powder was painted onto the commercial chips without encapsulation. The LEDs were driven by a Keithley 2400 source meter, and the emission spectra and intensity were recorded by a Photo Research SpectraScan PR655 photometer. For the device stability test, the LED was continuously powered by a Keithley 2400 source meter at a fixed current, and the initial brightness was set at about $5,000 \text{ cd m}^{-2}$. The device performance was monitored after a certain time interval.

Electroluminescence device performance measurement. The density–voltage and luminance–voltage characteristics and the electroluminescence spectra of the devices were collected by a Photo Research SpectraScan PR655 photometer and a Keithley 2400 source meter constant-current source. All the experiments were carried out at room temperature under ambient conditions in the dark.

Calculation and comparison of Huang–Rhys factors. In principle, the Huang–Rhys factor (S) reflects how strongly electrons couple to phonons and can be obtained by fitting the temperature-dependent full-width at half-maxima (FWHM) of photoluminescence peaks using the following equation⁵¹

$$\text{FWHM} = 2.36\sqrt{S\hbar\omega_{\text{phonon}}} \sqrt{\cot\left(\frac{\hbar\omega_{\text{phonon}}}{2k_B T}\right)} \quad (2)$$

where $\hbar\omega_{\text{phonon}}$ is the phonon frequency. For Cs₂AgInCl₆, S and $\hbar\omega_{\text{phonon}}$ are calculated as 38.7 and 20.1 meV, respectively, in good agreement with our simulation results (37 and 17.4 meV). Extended Data Table 1 lists the S values of a few representative compounds—we note that the Huang–Rhys factor of nanomaterials is generally higher than that of their bulk counterparts because of quantum confinement⁵². The Huang–Rhys factor of Cs₂AgInCl₆ is 38.7, which is larger than that of many common emitters, such as CdSe⁵³, ZnSe⁵⁴ and CsPbBr₃⁵⁵, indicating the easy formation of STEs in Cs₂AgInCl₆. For comparison, formation of STEs is also found in materials with high Huang–Rhys factors⁵⁶, such as Cs₃Sb₂I₉, Cs₃Bi₂I₉ and Rb₃Sb₂I₉. However, for efficient STE emission, S should not be overly large, because otherwise the excited-state energy would be dissipated by phonons, as is the case in Cs₂NaInCl₆. This is because S also influences photoluminescence

emission through the Franck–Condon factor, as described by equation (1). If we assume that the ground and excited states have similar phonon frequencies, the Franck–Condon factor (F , at zero temperature) can be simplified as

$$F = |\langle \chi_n | \chi_m \rangle|^2 = \frac{e^{-S} S^n}{n!} \quad (3)$$

The photoluminescence peak appears at $n \approx S$, so

$$F_{\text{max}} = \frac{e^{-S} S^S}{S!} \quad (4)$$

which is a monotonically decreasing function of S . Because the photoluminescence intensity is positively correlated with S , the larger the S , the smaller the radiative rate and the lower the emission efficiency. Thereby, the S value could potentially serve as the figure of merit for the design of efficient emission from STEs. The ideal value of the Huang–Rhys parameter should be intermediate for efficient STE emitters.

Mechanistic study of Bi³⁺ doping. Extended Data Fig. 6 provides information about the effect of Bi³⁺ incorporation on the PLQY improvement. For a Bi-doped Cs₂AgInCl₆ sample, the XRD measurement revealed smaller FWHM (from 0.058° to 0.034°) of the diffraction patterns, and the optical measurement demonstrated diminished sub-bandgap absorption after 400 nm and increased photoluminescence lifetime—from 2,971 ns (70%) to 5,989 ns (97%). Because an In³⁺ vacancy is a deep defect in Cs₂AgInCl₆, and isovalent doping helps to reduce vacancy defects in perovskite, we believe that Bi doping passivates defects and suppresses non-radiative recombination loss. Additionally, the theoretical simulation indicated that Bi doping introduces a shallow state right above the valence-band maximum and forms nanoelectronic domains in the matrix that concentrate holes. The holes finally relax to Ag sites through Bi 6s/Ag 4d orbital hybridization and lattice interaction, promoting exciton localization, just like I-doped AgBr for STE emission. Therefore, Bi doping improves crystal quality and promotes radiative recombination, enhancing the PLQY.

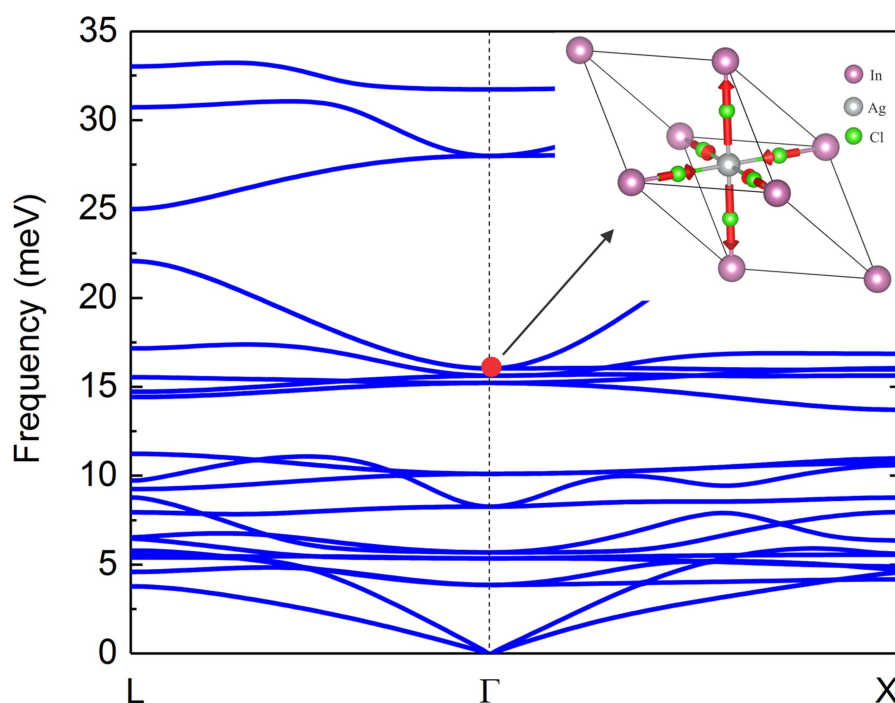
Code availability. The customized codes required for STE calculation with *cp2k* and the Python script used to calculate the Franck–Condon factors and luminescence spectrum are freely available at <https://github.com/wxiaom86>.

Data availability

The datasets analysed during the study are available from the corresponding authors upon request.

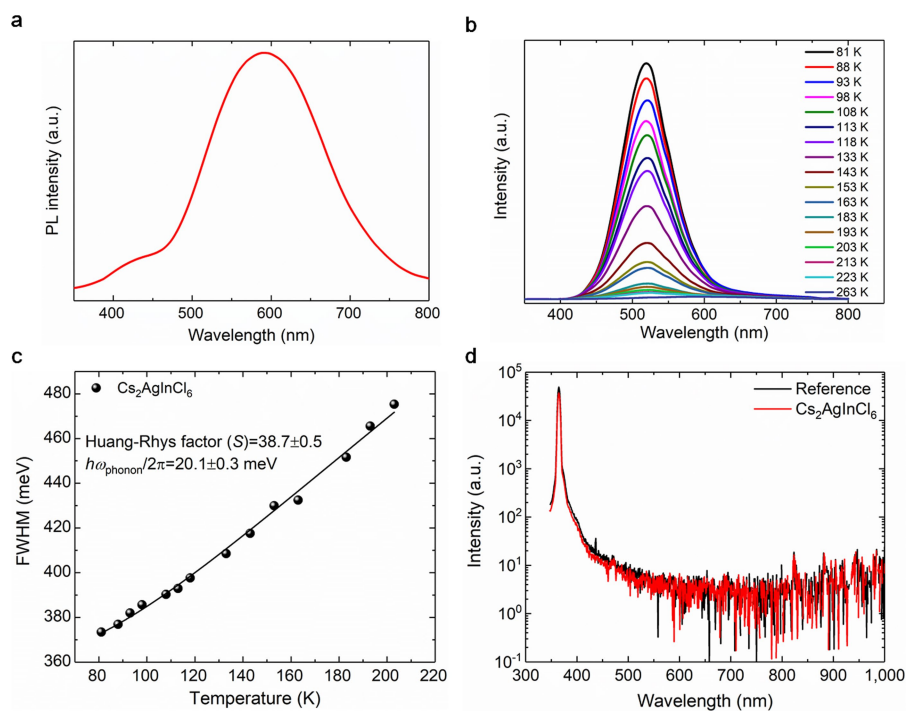
- Dai, X. et al. Solution-processed, high-performance light-emitting diodes based on quantum dots. *Nature* **515**, 96–99 (2014).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
- Kresse, G. & Furthmüller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane wave basis set. *Comput. Mater. Sci.* **6**, 15 (1996).
- Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
- Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982–9985 (1996).
- Perdew, J., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- Klimeš, J., Kaltak, M. & Kresse, G. Predictive GW calculations using plane waves and pseudopotentials. *Phys. Rev. B* **90**, 075125 (2014).
- Mostofi, A. A. et al. wannier90: a tool for obtaining maximally-localised Wannier functions. *Comput. Phys. Commun.* **178**, 685–699 (2008).
- Cappellini, G. et al. Model dielectric function for semiconductors. *Phys. Rev. B* **47**, 9892 (1993).
- Kowalczyk, T., Tsuchimochi, T., Chen, P. T., Top, L. & Van Voorhis, T. Excitation energies and Stokes shifts from a restricted open-shell Kohn–Sham approach. *J. Chem. Phys.* **138**, 164101 (2013).
- Filatov, M. & Shaik, S. A spin-restricted ensemble-referenced Kohn–Sham method and its application to diradicaloid situations. *Chem. Phys. Lett.* **304**, 429–437 (1999).
- Frank, I., Hutter, J., Marx, D. & Parrinello, M. Molecular dynamics in low-spin excited states. *J. Chem. Phys.* **108**, 4060–4069 (1998).
- Hutter, J., Iannuzzi, M., Schiffmann, F. & VandeVondele, J. Cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 15–25 (2014).
- VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).
- Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703–1710 (1996).
- Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).

47. VandeVondele, J. & Sprik, M. A molecular dynamics study of the hydroxyl radical in solution applying self-interaction-corrected density functional methods. *Phys. Chem. Chem. Phys.* **7**, 1363 (2005).
48. d'Avezac, M., Calandra, M. & Mauri, F. Density functional theory description of hole-trapping in SiO₂: a self-interaction-corrected approach. *Phys. Rev. B* **71**, 205210 (2005).
49. Alkauskas, A., Lyons, J. L., Steiauf, D. & Van De Walle, C. G. First-principles calculations of luminescence spectrum line shapes for defects in semiconductors: the example of GaN and ZnO. *Phys. Rev. Lett.* **109**, 267401 (2012).
50. Ruhoff, P. T. Recursion relations for multi-dimensional Franck–Condon overlap integrals. *Chem. Phys.* **186**, 355–374 (1994).
51. Stadler, W. et al. Optical investigations of defects in Cd_{1-x}Zn_xTe. *Phys. Rev. B* **51**, 10619 (1995).
52. Nandakumar, P. et al. Optical absorption and photoluminescence studies on CdS quantum dots in Nafion. *J. Appl. Phys.* **91**, 1509–1514 (2002).
53. Türcü, V. et al. Effect of random field fluctuations on excitonic transitions of individual CdSe quantum dots. *Phys. Rev. B* **61**, 9944 (2000).
54. Zhao, H. et al. Energy-dependent Huang–Rhys factor of free excitons. *Phys. Rev. B* **68**, 125309 (2003).
55. Lao, X. et al. Luminescence and thermal behaviors of free and trapped excitons in cesium lead halide perovskite nanosheets. *Nanoscale* **10**, 9949–9956 (2018).
56. McCall, K. M. et al. Strong electron–phonon coupling and self-trapped excitons in the defect halide perovskites A₃M₂I₉ (A=Cs, Rb; M=Bi, Sb). *Chem. Mater.* **29**, 4129–4145 (2017).
57. Leung, C. H. & Song, K. S. On the luminescence quenching of F centers in alkali halides. *Solid State Commun.* **33**, 907 (1980).
58. Mulazzi, E. & Terzi, N. Evaluation of the Huang–Rhys factor and the half-width of F-band in KCl and NaCl crystals. *J. Phys. Colloq.* **28**, 49–54 (1967).
59. Schulz, M. et al. Intensity dependent effects in silver chloride: bromine-bound exciton and biexciton states. *Phys. Status Solidi B* **177**, 201–212 (1993).
60. Andrews, L. J. et al. Thermal quenching of chromium photoluminescence in ordered perovskites. I. Temperature dependence of spectra and lifetimes. *Phys. Rev. B* **34**, 2735 (1986).



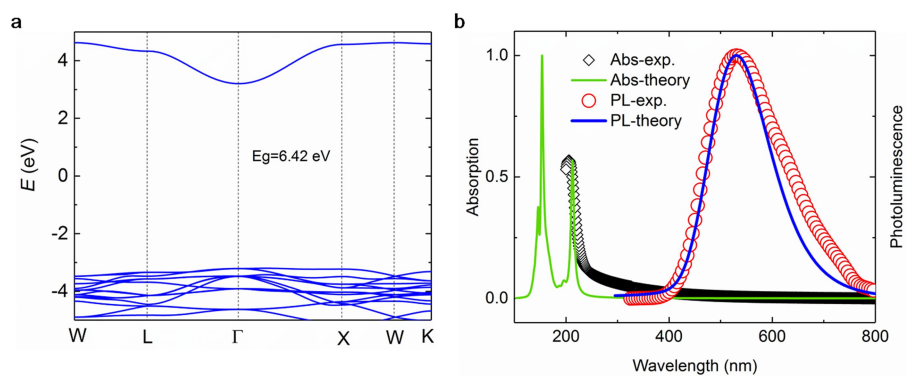
Extended Data Fig. 1 | Phonon band structure of $\text{Cs}_2\text{AgInCl}_6$ and the zone-centre Jahn–Teller phonon mode (inset). The phonon band structure was calculated by the finite-difference method with the supercell approach. The consistency of the displacement pattern of the phonon eigenvector with that of the lattice distortion during STE formation, as

well as the consistency of the phonon eigenfrequency with the phonon frequency fitted from the configuration coordinate diagram, confirm that the Jahn–Teller phonon mode coupled with the photoexcited excitons is responsible for the STE formation in $\text{Cs}_2\text{AgInCl}_6$.



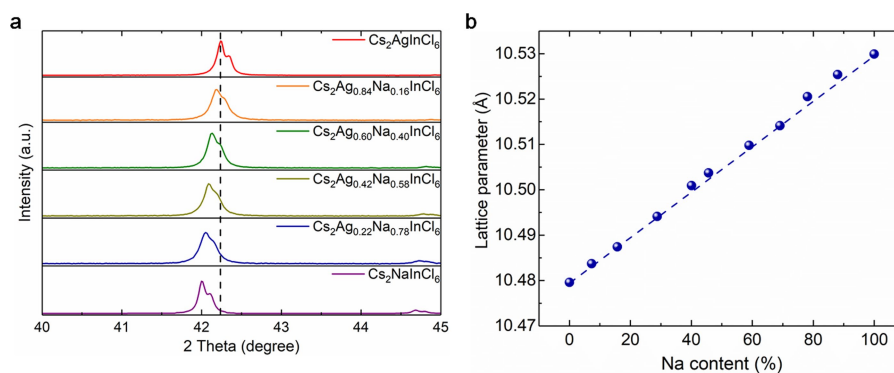
Extended Data Fig. 2 | Emission characterization of pure $\text{Cs}_2\text{AgInCl}_6$. **a**, The broad photoluminescence (PL) spectrum of $\text{Cs}_2\text{AgInCl}_6$ measured at room temperature. **b**, Temperature-dependent photoluminescence spectra of pure $\text{Cs}_2\text{AgInCl}_6$. **c**, Fitting results of the FWHM as a function

of temperature. We note that we used a relatively low-temperature region to avoid the influence of defect-assisted emission. **d**, The PLQY of $\text{Cs}_2\text{AgInCl}_6$. The reference was measured in an integrating sphere with a blank quartz plate.



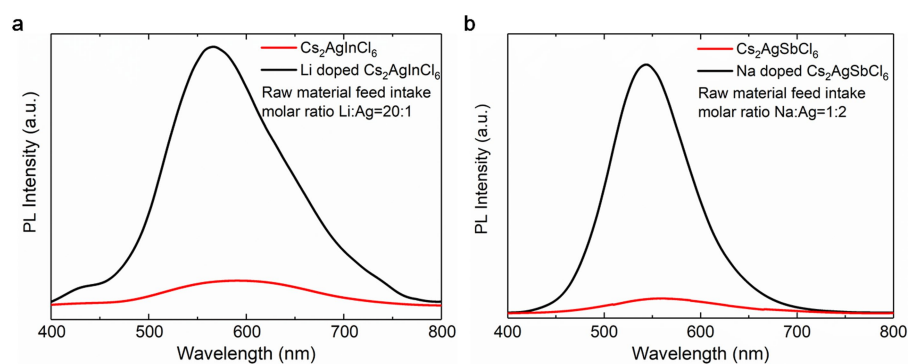
Extended Data Fig. 3 | Electronic and optical properties of $\text{Cs}_2\text{NaInCl}_6$. **a**, GW-calculated band structure. The GW bandgap is 6.42 eV. The lowest exciton, with a binding energy of 0.8 eV, is dark. The first bright exciton

has a binding energy of 0.44 eV. **b**, Calculated optical absorption ('Abs-theory') and photoluminescence ('PL-theory') spectra are compared with experimental results ('Abs-exp.' and 'PL-exp.').

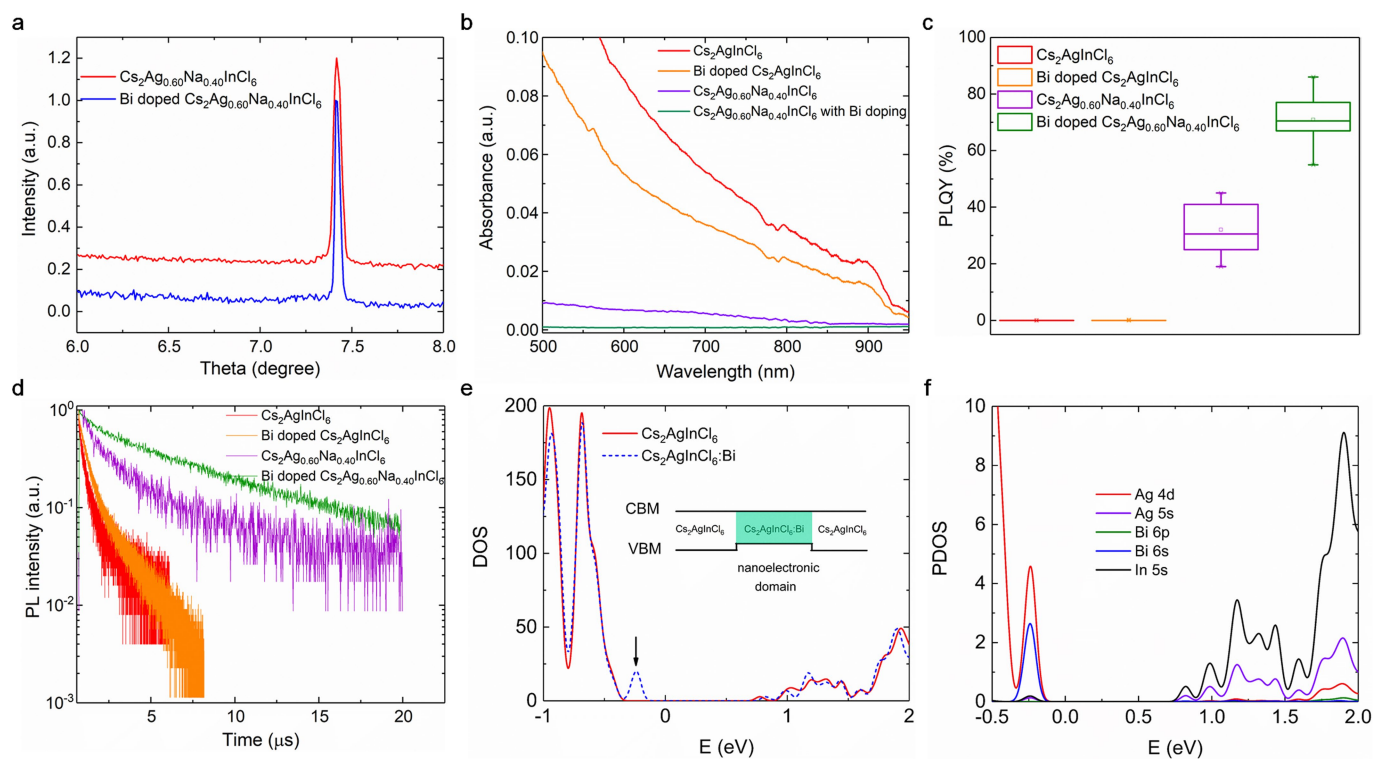


Extended Data Fig. 4 | Alloy behaviour of $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$. **a**, XRD patterns of $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$, shifted to lower degrees with increasing sodium substitution (theta, diffraction angle). **b**, Refined lattice parameter, plotted as a function of the nominal x in $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$, showing a linear increase with increased sodium substitution (see Supplementary

Fig. 3 for details of the characterization). We note that selected-area electron diffraction and scanning electron nanobeam diffraction analysis results (Supplementary Figs. 4, 5) suggest the existence of a microscopic super-lattice (Na/Ag ordering).



Extended Data Fig. 5 | Photoluminescence enhancement of doped double-perovskite powders. a, Photoluminescence spectra of pure $\text{Cs}_2\text{AgInCl}_6$ and Li-doped $\text{Cs}_2\text{AgInCl}_6$. **b,** Photoluminescence spectra of pure $\text{Cs}_2\text{AgSbCl}_6$ and Na-doped $\text{Cs}_2\text{AgSbCl}_6$.



Extended Data Fig. 6 | Characterization of the effect of Bi doping on $\text{Cs}_2\text{Ag}_x\text{Na}_{1-x}\text{InCl}_6$. **a**, High-resolution single-crystal XRD of the (111) peaks of $\text{Cs}_2\text{Ag}_{0.60}\text{Na}_{0.40}\text{InCl}_6$ with and without Bi doping. **b**, Absorption spectra of various materials with and without Bi doping for wavelengths of 500–950 nm. **c**, PLQY results. **d**, Photoluminescence lifetime. **e**, Comparison of the total density of states (DOS) between pure and

Bi-doped $\text{Cs}_2\text{AgInCl}_6$. The inset shows the band alignment of pure and Bi-doped $\text{Cs}_2\text{AgInCl}_6$. CBM, conduction band minimum; VBM, valence band maximum. The small shallow peak marked by an arrow is derived from the Bi 6s states, which hybridize with the Ag 4d states. **f**, Partial density of states (PDOS) of Bi-doped $\text{Cs}_2\text{AgInCl}_6$.

Extended Data Table 1 | Huang–Rhys factors

Compounds	Huang-Rhys factor
CdSe⁵³	1
ZnSe⁵⁴	0.3
CsPbBr₃⁵⁵	3.2
Cs₃Bi₂I₉⁵⁶	79.5
Cs₃Sb₂I₉⁵⁶	42.7
Rb₃Sb₂I₉⁵⁶	50.4
NaCl^{57,58}	42
AgCl:Br⁵⁹	22
Cs₂NaYCl₆⁶⁰	7.0
Cs₂NaInCl₆	80(ES)/188(GS)*
Cs₂AgInCl₆	38.7
Cs₂Ag_{0.60}Na_{0.40}InCl₆	40.9
Cs₂Ag_{0.16}Na_{0.84}InCl₆	51.0

The Huang–Rhys factors for CdSe, ZnSe, CsPbBr₃, NaCl and AgCl:Br are from the literature, Cs₂NaInCl₆ is a simulation result and the values for the other materials are obtained from the fitting results of the temperature-dependent FWHM of the photoluminescence data at a relatively low-temperature region (Extended Data Fig. 2c and Supplementary Fig. 1).

*Normally the Huang–Rhys factors of the ground state and the excited state are similar⁴⁹, as it is generally assumed that these states have the same phonon frequency. However, for Cs₂NaInCl₆ the difference is quite large: 80 at the excited state and 188 at the ground state.

Enhanced strength and ductility in a high-entropy alloy via ordered oxygen complexes

Zhifeng Lei^{1,10}, Xiongjun Liu^{1,10}, Yuan Wu^{1,10}, Hui Wang¹, Suihe Jiang¹, Shudao Wang¹, Xidong Hui¹, Yidong Wu¹, Baptiste Gault², Paraskevas Kontis², Dierk Raabe², Lin Gu³, Qinghua Zhang³, Houwen Chen⁴, Hongtao Wang⁵, Jiabin Liu⁶, Ke An⁷, Qiaoshi Zeng⁸, Tai-Gang Nieh⁹ & Zhaoping Lu^{1*}

Oxygen, one of the most abundant elements on Earth, often forms an undesired interstitial impurity or ceramic phase (such as an oxide particle) in metallic materials. Even when it adds strength, oxygen doping renders metals brittle^{1–3}. Here we show that oxygen can take the form of ordered oxygen complexes, a state in between oxide particles and frequently occurring random interstitials. Unlike traditional interstitial strengthening^{4,5}, such ordered interstitial complexes lead to unprecedented enhancement in both strength and ductility in compositionally complex solid solutions, the so-called high-entropy alloys (HEAs)^{6–10}. The tensile strength is enhanced (by 48.5 ± 1.8 per cent) and ductility is substantially improved (by 95.2 ± 8.1 per cent) when doping a model TiZrHfNb HEA with 2.0 atomic per cent oxygen, thus breaking the long-standing strength–ductility trade-off¹¹. The oxygen complexes are ordered nanoscale regions within the HEA characterized by (O, Zr, Ti)-rich atomic complexes whose formation is promoted by the existence of chemical short-range ordering among some of the substitutional matrix elements in the HEAs. Carbon has been reported to improve strength and ductility simultaneously in face-centred cubic HEAs¹², by lowering the stacking fault energy and increasing the lattice friction stress. By contrast, the ordered interstitial complexes described here change the dislocation shear mode from planar slip to wavy slip, and promote double cross-slip and thus dislocation multiplication through the formation of Frank–Read sources (a mechanism explaining the generation of multiple dislocations) during deformation. This ordered interstitial complex-mediated strain-hardening mechanism should be particularly useful in Ti-, Zr- and Hf-containing alloys, in which interstitial elements are highly undesirable owing to their embrittlement effects, and in alloys where tuning the stacking fault energy and exploiting athermal transformations¹³ do not lead to property enhancement. These results provide insight into the role of interstitial solid solutions and associated ordering strengthening mechanisms in metallic materials.

We studied the base alloy TiZrHfNb, the optimally oxygen-doped variant (TiZrHfNb)₉₈O₂ (hereafter denoted as O-2 HEA) and for comparison also an interstitial variant with 2.0 at% nitrogen (at%, atomic per cent), that is, (TiZrHfNb)₉₈N₂, hereafter referred to as N-2 HEA. Figure 1a shows the true tensile stress–strain curves of these three as-cast HEAs. A strong strengthening effect is observed for both the oxygen- and the nitrogen-doped HEAs: the yield strength σ_y increases from 0.75 ± 0.03 GPa for the base HEA to 1.11 ± 0.03 and 1.30 ± 0.02 GPa for the doped O-2 and N-2 HEAs, respectively. As expected from conventional interstitial strengthening, the ductility of the N-2 HEA is reduced, as indicated by the blue curve in Fig. 1a. Surprisingly, addition of 2.0 at% oxygen to the TiZrHfNb base HEA simultaneously improves

strength and ductility, as revealed by the red curve in Fig. 1a. The elongation has nearly doubled, increasing from $14.21\% \pm 1.09\%$ for the base HEA to $27.66\% \pm 1.13\%$ for the O-2 HEA, accompanied by a substantial work-hardening effect. We also observe a distinct yield point effect in the stress–strain curve of this specific alloy, which contrasts with the base and N-2 HEAs. This yield point phenomenon is similar to that observed also for many steels with solute carbon¹⁴, and results here from the interaction between oxygen and the dislocations. The O-2 HEA also exhibits a much higher work-hardening rate than the other two alloys (see inset in Fig. 1a), and the calculated hardening coefficient of the O-2 HEA is 0.124, which is much higher than that of the base HEA (0.042) and N-2 HEA (0.011), enabling the material's unexpected drastic increase in ductility. Figure 1b visualizes the anomalous interstitial strengthening behaviour of the O-2 HEA with respect to a number of established alloys^{15–23} and the N-2 HEA, revealing that its enormous increase in strength is not achieved at the expense of ductility. We note that the mechanical properties of the current TiZrHfNb are strongly dependent on the specific concentration of oxygen. Addition of more than 3.0 at% oxygen leads to deterioration of the mechanical properties, although oxides are still not formed.

To reveal the underlying mechanism of this anomalous interstitial solid-solution strengthening effect associated with oxygen in this material, we studied the underlying nanostructures in detail down to the atomic scale. Figure 2a shows the synchrotron high-energy X-ray diffraction (XRD) patterns of the base HEA as well as for the two alloy variants O-2 and N-2 HEAs, revealing that addition of either oxygen or nitrogen atoms into the base HEA has not changed its single-phase body-centred cubic (b.c.c.) structure. This observation is also confirmed by electron back-scattering diffraction mapping. The average grain size is similar for all three alloys (Fig. 2b), that is, the changes in mechanical behaviour are not due to the Hall–Petch or size effects. Figure 2c shows an aberration-corrected scanning transmission electron microscope high-angle annular dark field (STEM-HAADF) micrograph of the O-2 HEA with the incident electron beam aligned along the $[011]_{\text{b.c.c.}}$ zone axis of the grain selected. The Z-contrast image (where Z is the atomic number) is highly sensitive to local variations in the atomic number of the constituent elements²⁴, that is, light atoms exhibit dark contrast, whereas heavy atoms are imaged bright. The Z-contrast of the STEM-HAADF image reveals the existence of regions enriched in light atoms (that is, (Ti, Zr)-rich) and regions enriched in heavy atoms (that is, (Nb, Hf)-rich), highlighted by red and yellow squares, respectively, in Fig. 2d. This observation reveals the formation of compositionally short-range-ordered zones of metallic elements in the O-2 HEA. Similar zones also appear in the STEM images of the base and the N-2 HEAs (see Extended Data Fig. 1), confirming that such chemical short-range ordering among the metallic matrix elements is

¹Beijing Advanced Innovation Center for Materials Genome Engineering, State Key Laboratory for Advanced Metals and Materials, University of Science and Technology Beijing, Beijing, China.

²Department of Microstructure Physics and Alloy Design, Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany. ³Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China. ⁴College of Materials Science and Engineering, Chongqing University, Chongqing, China. ⁵Institute of Applied Mechanics, Zhejiang University, Hangzhou, China. ⁶School of Materials Science and Engineering, Zhejiang University, Hangzhou, China. ⁷Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, TN, USA. ⁸Center for High Pressure Science and Technology Advanced Research, Pudong, Shanghai, China. ⁹Department of Materials Science and Engineering, University of Tennessee, Oak Ridge, TN, USA. ¹⁰These authors contributed equally: Zhifeng Lei, Xiongjun Liu, Yuan Wu. *e-mail: luzp@ustb.edu.cn

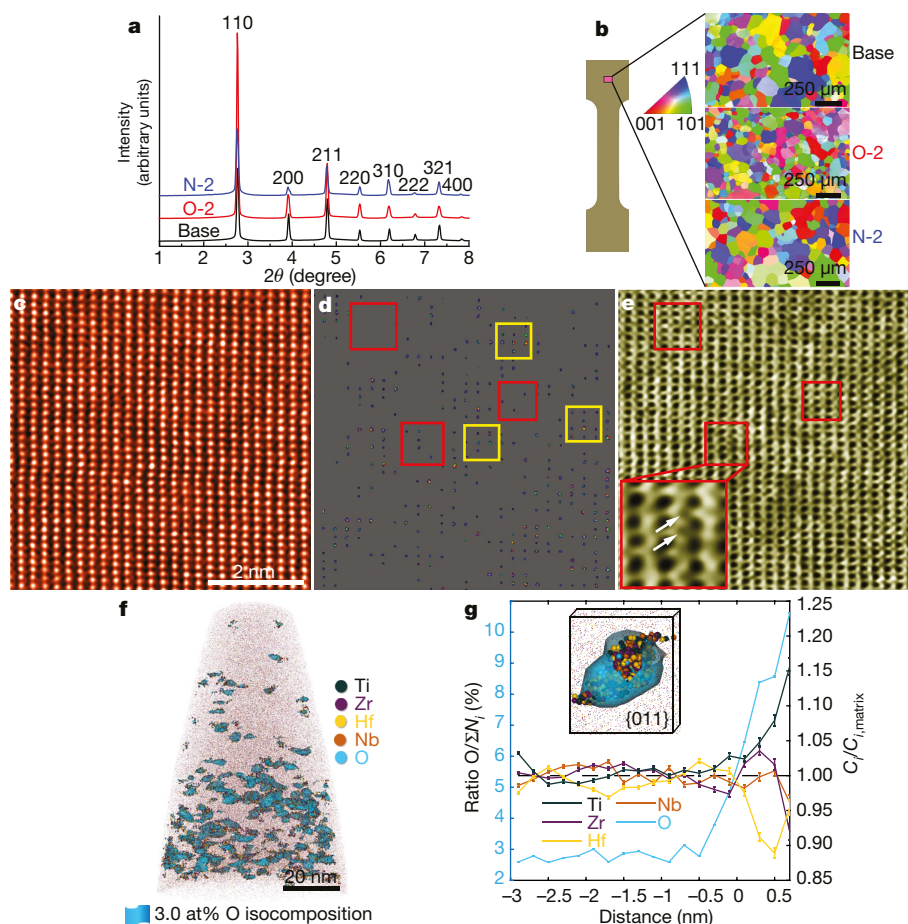
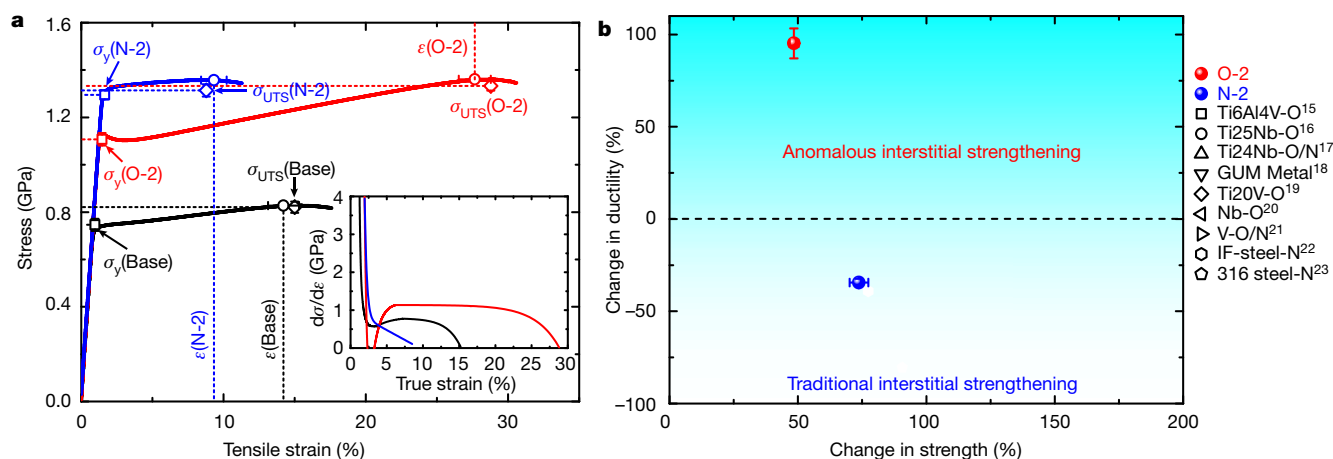


Fig. 2 | Microscopic structure. **a**, **b**, Synchrotron high-energy XRD and the corresponding electron back-scattering diffraction patterns of the as-cast equiatomic TiZrHfNb and the interstitially doped solid-solution HEAs. All the as-cast HEAs have single b.c.c. lattice structure. **c–e**, STEM-HAADF images for the [011]_{b.c.c.} crystal axis with differently adjusted contrast to reveal the existence of chemical short-range ordering in the O-2 HEA (TiZrHfNb)₉₈O₂, and the corresponding STEM-ABF image that reveals the ordered oxygen complexes (OOCs). Red squares represent the Zr/Ti-rich regions and yellow squares indicate the Hf/Nb-rich regions. The inset in **e** is an enlarged view of the OOCs, with the white arrows indicating the positions of the oxygen atom columns. **f**, Atom probe tomography image of the O-2 HEA. The threshold for the iso-composition surface is 3.0 at% O. **g**, O composition profile as a function of the distance to the interface for a selection of particles (left axis) and evolution of the composition of the main constituents relative to their respective matrix composition (right axis). The inset shows a close-up of one such OOC, along with the {011} atomic plane imaged within the reconstruction. N_i is the number of the i th atom, while C_i and $C_{i,matrix}$ are the concentrations of the i th atom in the OOCs and in the matrix, respectively. The error bars are standard deviations of the mean.

tomography three-dimensional reconstruction from the analysis of a specimen from the O-2 HEA. The threshold for the iso-composition surface is 3.0 at% O, highlighting the presence of OOCs. **g**, O composition profile as a function of the distance to the interface for a selection of particles (left axis) and evolution of the composition of the main constituents relative to their respective matrix composition (right axis). The inset shows a close-up of one such OOC, along with the {011} atomic plane imaged within the reconstruction. N_i is the number of the i th atom, while C_i and $C_{i,matrix}$ are the concentrations of the i th atom in the OOCs and in the matrix, respectively. The error bars are standard deviations of the mean.

an intrinsic feature of these HEAs. The aberration-corrected STEM-ABF (annular bright field) image (Fig. 2e) also shows that oxygen occupies with similar frequency both the tetrahedral and the octahedral interstitial sites in the b.c.c. lattice. This surprising observation is in good agreement with simulations based on first-principles methods (see Extended Data Fig. 2). More interestingly, statistical analysis of the STEM-HAADF and the corresponding ABF images (Fig. 2d, e) demonstrates that oxygen tends to prefer interstitial positions adjacent to light-atom-rich (for example, Zr and Ti) lattice sites and form ordered oxygen complexes (OOCs) with a length scale of 1–3 nm and spacing of 2–4 nm, as indicated by the red squares in Fig. 2d, e. By contrast, no such selective ordering phenomenon was observed in the N-2 HEA, as shown in Extended Data Fig. 1d–f. Further atom probe tomography measurements of the O-2 HEA are shown in Fig. 2f, g. A blue surface encompassing regions of the point cloud containing more than 3.0 at% O is superimposed on the point cloud, as shown in Fig. 2f, revealing the presence of O-rich clusters within the data, corresponding to the OOCs. To ensure that these clusters are not related to random fluctuations in the solid solution, a randomized dataset was generated and an iso-composition surface with the same threshold reveals no such clusters (see Extended Data Fig. 3), further confirming that oxygen locally assumes an ordered state in the matrix. An average composition profile as a function of the distance to the isosurface (see Methods) was calculated from 12 particles of similar size. The resulting O profile is plotted in Fig. 2g, along with the composition of the main constituents relative to their composition away from the particle, that is, in the matrix, which reveals an excess of Ti and a slight enrichment in Zr within the O-rich clusters. These observations are consistent with the ABF images (Fig. 2d, e) observed by STEM-HAADF in the O-2 HEA.

To understand better the statistics of this occupation difference between oxygen and nitrogen, we conducted internal-friction measurements on both interstitial alloy variants O-2 and N-2 HEAs; see Extended Data Fig. 4. We found that the O-2 HEA shows an extra peak at low temperatures compared to the N-2 HEA. This result confirms that oxygen assumes an additional structural state, namely, in the form of OOCs, which is characterized by an individual trapping barrier, a state not observed for nitrogen.

According to a solid-solution strengthening model proposed by Fleischer²⁵, we have calculated the tetragonal distortion $\Delta\epsilon$ for the O-2 and N-2 HEAs and confirmed that their hardening mechanism is indeed of interstitial nature (see Methods). This elastic analysis explains the influence of both interstitial solute oxygen and nitrogen on the strength of the HEA. Yet, unlike traditional interstitial strengthening, which often embrittles alloys, the presence of oxygen simultaneously increases not only the strength but also the ductility in the current b.c.c. TiZrHfNb HEA. To understand this phenomenon associated with the presence of oxygen, ex situ XRD and in situ neutron diffraction loading experiments of all three alloys were conducted (Extended Data Fig. 5). We found that no phase transformation during deformation occurred in any of the three alloys. Also, ex situ transmission electron microscopy (TEM) and in situ TEM mechanical testing experiments confirm that deformation of the three alloys works via dislocation glide (Extended Data Fig. 6 and Supplementary Videos).

In general, dislocations are stored and arranged in ordered patterns during plastic deformation. Their motion is also patterned with two prevalent types of mesoscopic deformation modes: planar slip and wavy slip²⁶. Continuous planar propagation of dislocations often prevails in face-centred cubic metals^{27–29}, whereas plasticity in b.c.c. metals is strongly influenced by frequent cross-slip owing to the similarly close-packed $\{110\}$ and $\{112\}$ slip planes, resulting in wavy slip patterns³⁰. To study these dislocation pattern features in more detail in the current alloys, we conducted high-resolution aberration-corrected STEM characterization of the pre-strained specimens (Fig. 3). Coplanar dislocation arrays resulting from planar slip are observed in the 8% pre-strained base alloy (see Fig. 3a). By contrast, well defined dipolar dislocation walls are developed in the 8% deformed O-2 HEA. Such substructures are usually caused by cross-slip of screw dislocations³¹

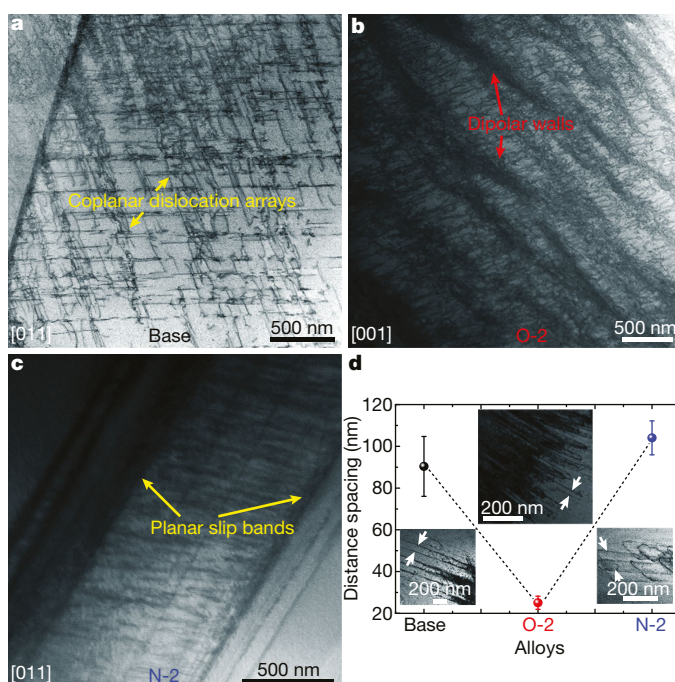


Fig. 3 | Deformation mode. **a**, STEM image of the TiZrHfNb base HEA at 8% tensile strain (the yellow arrows indicate the coplanar dislocation arrays). **b**, STEM image of O-2 HEA at 8% tensile strain (the red arrows indicate the dipolar walls). **c**, STEM image of N-2 HEA at 8% tensile strain (the yellow arrows indicate the planar slip bands). Typical planar slip is observed in the base HEA and in the nitrogen-doped alloy variant N-2 HEA. However, wavy slip dominates deformation of the oxygen-doped variant O-2 HEA, suggesting that oxygen addition leads to a plastic deformation mode dominated by wavy slip. The beam direction in **a** and **c** is $[011]$ while that in **b** is $[001]$. **d**, Dislocation spacing of the TiZrHfNb base HEA and of the interstitially doped variants O-2 and N-2 HEAs probed during in situ TEM tensile experiments. The white arrows represent the dislocation spacing. The average dislocation spacing in the O-2 HEA (25.06 ± 3.15 nm) is much smaller than that in the base HEA (90.36 ± 14.32 nm) and in the N-2 HEA (104.06 ± 8.14 nm). The error bars are standard deviations of the mean.

(Fig. 3b), indicating that the plastic deformation of the O-2 HEA is dominated by wavy slip. This observation suggests that addition of oxygen facilitates cross-slip, leading to the plastic deformation mode, which is characterized not by planar slip but by wavy slip. However, for the N-2 HEA, well developed planar and banded dislocation substructures are observed (see Fig. 3c), revealing that the deformation mode is similar to that of the base HEA.

We conducted a more detailed analysis to clarify the differences in the dislocation substructures and slip band morphologies among the three HEAs formed during deformation. The experiments show that the O-2 HEA is characterized by frequent dislocation cross-slip (Extended Data Fig. 7). We also found that both the dislocation density and their velocity increase dramatically in the O-2 HEA but decrease in the N-2 variant (see Fig. 3d and Supplementary Videos 1–3). These fundamentally different dislocation multiplication, glide and patterning features further indicate that the enhanced ductility of the O-2 HEA results from facilitated cross-slip and the associated promotion of dislocation nucleation and propagation.

The key question from the substructure analysis is why the addition of interstitial oxygen not only greatly changes the dislocation patterning, motion and multiplication, but enhances the alloy's work-hardening capacity and thereby its ductility. As elaborated above, the main difference between oxygen and nitrogen in this alloy class is that oxygen occupies the interstitial sites in the Zr and/or Ti-enriched clusters in an orderly manner. By contrast, no ordered nitrogen-containing complexes are present. The nanoscale oxygen-containing complexes

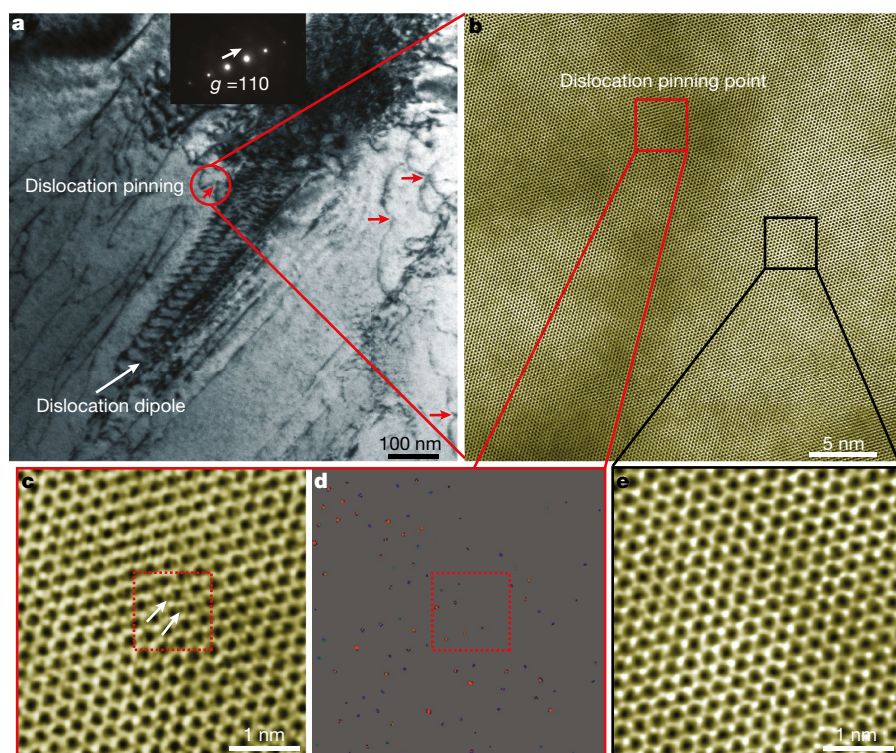


Fig. 4 | Intrinsic mechanism. **a**, Dislocations in the 8% strained O-2 HEA, imaged under $\{111\}$ -type diffraction conditions. Dislocation pinning at OOCs is observed, which suppresses dislocation motion (red arrows). Also, dislocation dipoles produced by dislocation cross-slip are found (white arrow). A dislocation pinning point (red circle) at such an ordered complex was chosen for further STEM characterization in **b** and **d**. **g** indicates the direction of the diffraction vector. **b**, Aberration-corrected STEM-ABF image of the local atomic structure near the dislocation pinning point. Atomic structure analysis of the regions at the dislocation pinning point (red square) and away from the pinning point (black square)

were conducted. **c**, Aberration-corrected STEM-ABF image of the local atomic structure at the pinning point. The white arrows point to the oxygen atom columns around the pinned dislocation. Interstitial oxygen atoms are clearly seen in the red dotted square. **d**, Corresponding STEM-HAADF image for the $[111]_{b.c.c.}$ crystal axis with differently adjusted contrast to reveal that the pinning effect is induced by OOCs. The red dotted zone indicates the (O, Zr, Ti)-rich region, that is, the OOC. **e**, Aberration-corrected STEM-ABF image of the local atomic structure away from the pinning point, where no similar ordered interstitial complexes were observed.

severely distort the local lattice, leading to a large strain field around them. During deformation, these OOCs interact with dislocations, that is, via pinning and promotion of dislocation double cross-slip, as revealed by STEM analyses (Extended Data Fig. 8).

To examine the interaction between OOCs and dislocations, atomic-scale structure characterization of the 8% pre-tensioned O-2 HEA was conducted by aberration-corrected STEM (Fig. 4). Dislocation dipoles produced by dislocation cross-slip and the distinct pinning effect (red arrows) of the pre-strained O-2 HEA are clearly observed (see Fig. 4a). Some of these configurations were chosen for atomic structure observation using aberration-corrected STEM-ABF imaging (Fig. 4b). Two zones, that is, the region at the dislocation pinning point (red square) and away from the point (black square), were picked out for further local atomic structure characterization, as shown in Fig. 4c–e. In the region containing the pinned dislocation, interstitial oxygen atoms are clearly observed (white arrows in Fig. 4c). The corresponding STEM-HAADF image for the $[111]_{b.c.c.}$ crystal axis with adjusted contrast reveals oxygen atoms adjacent to Zr/Ti atoms (red dotted square in Fig. 4d). All these observations prove that the pinning of dislocations is indeed caused by OOCs. For the region away from the pinning point, no such OOCs are observed (see Fig. 4e). These results indicate that the OOCs have a twofold role during deformation: first, as dislocation pinning points similar to very small precipitates, and second, through the homogenization of the plastic flow by switching the slip mode from planar to wavy slip.

From the above analyses, the underlying deformation mechanisms responsible for the unprecedented interstitial strengthening in the O-2 HEA are schematically illustrated in Extended Data Fig. 9. In the base and N-2 HEAs, conventional substitutional clustering and

compositional short-range ordering of the metallic constituents promote planar dislocation slip. Plastic deformation in these alloys is thus topologically confined and localized to a few single-slip planes. Large dislocation densities thus assemble along the same glide planes, leading to in-plane softening and high pile-up stresses, promoting damage initiation²⁹. In the O-2 HEA variant, however, the OOCs act on dislocations with features that lie between the effects known from conventional interstitial strengthening and those known from nano-precipitates (Extended Data Fig. 9a). During the very first stages of plastic deformation, planar slip still prevails (Extended Data Fig. 9b), but once the dislocations encounter the severely distorted, interstitial-enriched OOCs, cross-slip is promoted owing to their strong pinning effects (Extended Data Fig. 9c), resulting in massive dislocation multiplication (Extended Data Fig. 9d). Upon reaching a higher dislocation density and thus higher stress levels, dislocations also start to pass through these OOCs, promoting planar slips again until other OOCs are encountered. The interplay of sequential pinning, cutting and cross-slip leads to the homogenization of the dislocation substructure, which, on the one hand, avoids local stress peaks observed in purely planar dislocation arrays, and, on the other hand, promotes high multiplication rates of dislocation via double cross-slip and the formation of new Frank–Read sources. These features lead to high work hardening, as demonstrated in Fig. 1a. Subsequently, more and more dislocations are pinned by OOCs and dipolar walls emerge as the strain increases (Extended Data Fig. 9e), which further promotes work hardening and delays the onset of necking, eventually leading to higher ductility. For the N-2 HEA variant, we observe no such complex–dislocation dynamics and structures, that is, conventional planar dislocation slip prevails. As a result, multiplication and propagation of dislocations are limited, leading to modest ductility.

In summary, the current findings not only show how the strength–ductility trade-off can be successfully overcome for a class of HEAs, but also we elucidate a completely new type of strain-hardening mechanism based on ordered interstitial complexes. This effect enables an excellent balance between dislocation pinning, multiplication and substructure homogenization, and thereby leads to a high strain-hardening reserve and an increase in both strength and ductility. The current TiZrHfNb material in its current alloy design stage is not yet suitable for immediate utilization in high-temperature applications because of oxidation problems. Alloying with antioxidant elements, such as Al, Si and Cr, could improve the oxidation resistance of these HEAs, as has been successfully demonstrated in other types of HEAs³². We suggest that this type of ordered interstitial complex strengthening mechanism is also applicable to a wide range of other alloy classes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0685-y>.

Received: 7 March; Accepted: 14 September 2018;

Published online 14 November 2018.

- Conrad, H. Effect of interstitial solutes on the strength and ductility of titanium. *Prog. Mater. Sci.* **26**, 123–403 (1981).
- Tyson, W. R. Strengthening of hcp Zr, Ti and Hf by interstitial solutes—a review. *Can. Metall. Q.* **6**, 301–332 (1967).
- Mouawad, B., Boulmat, X., Fabrége, D., Perez, M. & De Carlan, Y. Tailoring the microstructure and the mechanical properties of ultrafine grained high strength ferritic steels by powder metallurgy. *J. Nucl. Mater.* **465**, 54–62 (2015).
- Wei, Q. et al. Influence of oxygen content on microstructure and mechanical properties of Ti–Nb–Ta–Zr alloy. *Mater. Des.* **32**, 2934–2939 (2011).
- Ando, T., Nakashima, K., Tsuchiyama, T. & Takaki, S. Microstructure and mechanical properties of a high nitrogen titanium alloy. *Mater. Sci. Eng. A* **486**, 228–234 (2008).
- Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375/377**, 213–218 (2004).
- Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
- Zhang, Y. et al. Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93 (2014).
- Li, Z. M., Pradeep, K. G., Deng, Y., Raabe, D. & Tasan, C. C. Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off. *Nature* **534**, 227–230 (2016).
- Gludovatz, B. et al. A fracture resistant high entropy alloy for cryogenic applications. *Science* **345**, 1153–1158 (2014).
- Ritchie, R. O. The conflicts between strength and toughness. *Nat. Mater.* **10**, 817–822 (2011).
- Wang, Z. et al. The effect of interstitial carbon on the mechanical properties and dislocation substructure evolution in Fe_{40.4}Ni_{11.3}Mn_{34.8}Al_{7.5}Cr₆ high entropy alloys. *Acta Mater.* **120**, 228–239 (2016).
- Zhu, Y. T. & Liao, X. Nanostructured metals: retaining ductility. *Nat. Mater.* **3**, 351–352 (2004).
- Cottrell, A. H. & Bilby, B. A. Dislocation theory of yielding and strain ageing of iron. *Proc. Phys. Soc. A* **62**, 49 (1949).
- Oh, J. M. et al. Oxygen effects on the mechanical properties and lattice strain of Ti and Ti–6Al–4V. *Met. Mater. Int.* **17**, 733–736 (2011).
- Yin, F., Iwasaki, S., Ping, D. & Nagai, K. Snoek-type high-damping alloys realized in β -Ti alloys with high oxygen solid solution. *Adv. Mater.* **18**, 1541–1544 (2006).
- Ramarolahy, A. et al. Microstructure and mechanical behavior of superelastic Ti–24Nb–0.5O and Ti–24Nb–0.5N biomedical alloys. *J. Mech. Behav. Biomed. Mater.* **9**, 83–90 (2012).
- Besse, M., Castany, P. & Gloriant, T. Mechanisms of deformation in gum metal TNTZ–O and TNTZ titanium alloys: a comparative study on the oxygen influence. *Acta Mater.* **59**, 5982–5988 (2011).
- Wang, X., Li, L., Xing, H., Ou, P. & Sun, J. Role of oxygen in stress-induced ω phase transformation and {332}<113> mechanical twinning in β Ti–20V alloy. *Scr. Mater.* **96**, 37–40 (2015).
- Sankar, M., Baligidad, R. G. & Gokhale, A. A. Effect of oxygen on microstructure and mechanical properties of niobium. *Mater. Sci. Eng. A* **569**, 132–136 (2013).
- Jo, M. G., Madakashira, P. P., Suh, J. Y. & Han, H. N. Effect of oxygen and nitrogen on microstructure and mechanical properties of vanadium. *Mater. Sci. Eng. A* **675**, 92–98 (2016).
- Shen, Y. Z., Oh, K. H. & Lee, D. N. Nitrogen strengthening of interstitial-free steel by nitriding in potassium nitrate salt bath. *Mater. Sci. Eng. A* **434**, 314–318 (2006).
- Talha, M., Behera, C. K. & Sinha, O. P. Effect of nitrogen and cold working on structural and mechanical behavior of Ni-free nitrogen containing austenitic stainless steels for biomedical applications. *Mater. Sci. Eng. C* **47**, 196–203 (2015).
- Pennycook, S. J., Rafferty, B. & Nellist, P. D. Z-contrast imaging in an aberration-corrected scanning transmission electron microscope. *Microsc. Microanal.* **6**, 343–352 (2000).
- Fleischer, R. L. Solution hardening by tetragonal distortions: application to irradiation hardening in FCC crystals. *Acta Metall.* **10**, 835–842 (1962).
- Hong, S. I. & Laird, C. Mechanisms of slip mode modification in F.C.C. solid solutions. *Acta Metall.* **38**, 1581–1594 (1990).
- Yao, M. J., Pradeep, K. G., Tasan, C. C. & Raabe, D. A novel, single phase, non-equiatomic FeMnNiCoCr high-entropy alloy with exceptional phase stability and tensile ductility. *Scr. Mater.* **82**, 5–8 (2013).
- Yoo, J. D. & Park, K. T. Microband-induced plasticity in a high Mn–Al–C light steel. *Mater. Sci. Eng. A* **496**, 417–424 (2008).
- Gerold, V. & Karnthaler, H. P. On the origin of planar slip in f.c.c. alloys. *Acta Metall.* **37**, 2177–2183 (1989).
- Nabarro, F. R. & Duesbery, M. S. *Dislocations in Solids* Vol. 11 (Elsevier, Amsterdam, 2002).
- Mughrabi, H., Ackermann, F. & Herz, K. *Fatigue Mechanisms ASTM STP 675 69* (American Society for Testing Materials, Philadelphia, 1979).
- Liu, C. M., Wang, H. M., Zhang, S. Q., Tang, H. B. & Zhang, A. L. Microstructure and oxidation behavior of new refractory high entropy alloys J. *Alloys Compd.* **583**, 162–169 (2014).

Acknowledgements This research was supported by National Natural Science Foundation of China (grant numbers 51671018, 11790293, 51871016, 51531001 and 51671021), the 111 Project (grant number B07003), the Program for Changjiang Scholars and Innovative Research Team in University of China (grant number IRT_14R05) and the Projects of SKLMM-USTB (grant numbers 2018Z-01 and 2018Z-19). Yuan W. acknowledges financial support from the Top-Notch Young Talents Program. Yuan W. and Hui W. acknowledges financial support from the Fundamental Research Funds for the Central Universities. We thank F. Zhang at the University of Science and Technology Beijing for help with synchrotron XRD. We also thank H. L. Huang at the University of Science and Technology Beijing and L. Qi and X. J. Zhao at the Chongqing University for help with TEM/STEM characterization and discussion.

Author contributions Z. Lu designed the study. Z. Lei, X.L., Yuan W., Hui W., S.J., S.W., X.H. and Yidong W. carried out the main experiments. Z. Lei, X.L., Yuan W., Z. Lu, B.G. and D.R. analysed the data and wrote the main draft of the paper. L.G., Q.Z., H.C., Hongtao W. and J.L. conducted the TEM and STEM characterizations. B.G., P.K. and D.R. prepared the atom probe tomography specimens, processed the data and interpreted the results. K.A. conducted the neutron diffraction. Q.Z. conducted the synchrotron XRD. All authors contributed to the discussion of the results, and commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0685-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0685-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to Z.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Material preparation. HEAs have desirable properties compared to traditional alloys with one major component^{33–36}. Among them, HEAs consisting of metallic elements with very high melting points (>1,900 K) have attracted considerable attention owing to features such as good softening resistance at elevated temperatures and slow diffusion kinetics^{37–39}. Here, alloy ingots with a nominal composition of TiZrHfNb, (TiZrHfNb)₉₈O₂ and (TiZrHfNb)₉₈N₂ (at%) were prepared by arc-melting a mixture of pure metals (purity >99.9 wt%), TiN (99.9 wt%) and TiO₂ (99.9 wt%) in a Ti-gettered high-purity argon atmosphere. The ingots were remelted at least eight times to ensure chemical homogeneity. Melted alloys were eventually drop-cast into a water-cooled copper mould with dimensions of 10 mm × 10 mm × 60 mm.

XRD. Synchrotron XRD patterns were obtained from each of the alloys at the Advanced Photon Source at Argonne National Laboratory, USA. Two-dimensional diffraction patterns were collected in transmission geometry using a PerkinElmer α-Si flat-panel large-area detector at the 11-ID-C beam line. The wavelength of the X-ray was about 0.117418 Å, and the beam size was 500 × 500 μm². One-dimensional XRD patterns were obtained by integrating the two-dimensional patterns with Fit 2d software (<http://www.esrf.eu/computing/scientific/FIT2D/>). Phase identification of the as-cast alloys and the deformed alloys were also conducted by XRD using Cu Kα radiation (XRD model MXP21VAHF).

SEM. Microstructure and morphology were characterized by a Zeiss Supra 55 field emission scanning electron microscope equipped with an AZtecHKL electron back-scattering diffraction system. The electron back-scattering diffraction specimens were initially polished with 2,000-grit SiC paper and subsequently electrochemically polished using a 6% perchloric acid + 30% *n*-butyl alcohol + 64% methyl alcohol solution at a direct voltage of 30 V at room temperature (298 K). The tensile samples used for surface morphology observation were also electrochemically polished.

Atom probe tomography. Atom probe tomography analyses were carried out for the base, O-2 and N-2 HEAs in a Cameca LEAP 3000X HR under ultrahigh vacuum of approximately 2.5 × 10^{−11} torr, at a specimen temperature of 80 K, a target evaporation rate of 3 ions for 1,000 pulses on average in high-voltage pulsing mode at 15% pulse fraction. Atom probe tomography specimens were prepared by focused ion beam milling on a dual-beam FEI Helios 600 using the protocol outlined in ref. 40. The CAMECA integrated visualization and analysis software IVAS 3.6.8 was used for data processing and three-dimensional atomic reconstruction.

Mechanical property measurements. Room-temperature tensile properties were evaluated using a CMT4105 universal electronic tensile testing machine with an initial strain rate of 2.0 × 10^{−4} s^{−1}. Dogbone-shaped tensile samples with a cross section of 1.0 × 3.0 mm² and a gauge length of 20 mm were cut using electrical discharging. At least 5 samples were tested for each composition. Hardness measurements were conducted using a Vickers hardness tester (430SVD) with a load of 5 N for 15 s, and for each specimen, at least 15 indents were measured to obtain an average value. Nanoindentation tests were performed using the Nanoindenter XP system equipped with a spherical indenter with a radius of 1 μm, and at least 100 indents were measured to obtain an average value of modulus. The samples for hardness measurements and nanoindentation tests were prepared by electrochemical polishing.

In situ neutron diffraction measurements. In situ neutron diffraction investigations upon tensile loading were conducted using a MTS load-frame on the VULCAN diffractometer, at the Spallation Neutron Source of Oak Ridge National Laboratory. VULCAN is equipped with two detector banks positioned at ±90° diffraction angles, which are designated banks 1 and 2, respectively. An incident neutron beam of 3 mm high and 3 mm wide, together with a pair of 5 mm radial collimators, were used to define the sampling volume during the diffraction experiments. The neutron diffraction data were reduced and then analysed by single peak fitting using the VDRIVE program⁴¹.

TEM. Microstructure of the as-cast and fractured specimens was characterized by high-resolution TEM with a JEM-2010. In situ TEM observations were conducted with a JEM-2100 TEM equipped with a tensile loading stage. An aberration-corrected FEI Titan G260–300 kV S/TEM was used to analyse the atomic structure and morphology of the 8% pre-strained samples. The TEM specimens were first mechanically ground to 50 μm thickness and then twin-jet electropolished using 6% perchloric acid + 30% *n*-butyl alcohol + 64% methyl alcohol solution.

Internal-friction measurements. Beam-shaped samples with dimensions of 1 mm × 2 mm × (35–55) mm were used for damping-capacity measurement on a multifunctional internal-friction device (MFP-1000 at the Institute of Solid State Physics, Chinese Academy of Sciences) at low frequencies of 0.5 Hz, 1.0 Hz, 2.0 Hz and 4.0 Hz over a temperature range from 300 K to 1,000 K under continuous heating in vacuum. All samples were polished using a 2,000-grit SiC paper to eliminate surface scratches. A heating rate of 2 K min^{−1} and a forced vibration with a strain amplitude of 2 × 10^{−5} were applied for the damping measurements. Damping data were collected using a fully automated system that measured the angular velocity of

the pendulum around the equilibrium position⁴². The background was subtracted according to the following equation

$$Q_b^{-1} = A + B \exp\left(\frac{-C}{kT}\right) \quad (1)$$

where Q_b^{-1} is the energy dissipation coefficient of the background, and A, B and C are the parameters to be determined after optimization of the χ^2 function⁴³.

Theoretical calculations and modelling. The first-principles calculations were conducted using the density functional theory (DFT)-based Vienna ab initio simulation package (VASP)⁴⁴ using projector augmented waves (PAW)⁴⁵ and the generalized gradient approximation of Perdew–Burke–Ernzerhof (GGA-PBE)⁴⁶ for the exchange correlation functional. A 72-atom b.c.c.-like special quasirandom structure⁴⁷ supercell was constructed to model the quaternary TiZrHfNb HEA which served as the base reference alloy state. To investigate the occupation probability of oxygen or nitrogen in the HEA, we placed one interstitial atom of oxygen or nitrogen at the octahedral or tetrahedral interstitial sites, respectively, in the supercell and calculated the system energy for the case of the oxygen/nitrogen at these different interstitial sites. The occupation probability was then estimated by the statistical distribution of system energy as shown in Extended Data Fig. 2. Structural relaxations and final static calculations were performed to obtain stable structures and more accurate energy values using the conjugate gradient method⁴⁸ and the linear tetrahedron method including Blöchl corrections⁴⁹, respectively. The energy cutoff on the wave function is taken as 450 eV, which is 1.3 times higher than the default value for all elements in this work. A 2 × 3 × 3 Monkhorst-Pack-ENREF-250 *k*-point mesh was used to sample the Brillouin zone. The energy convergence criterion for the electronic self-consistency was chosen as 10^{−6} eV per atom.

Theoretical calculations of interstitial strengthening. Interstitials in b.c.c. metals usually produce a tetragonal distortion of the lattice. The interaction between screw dislocations and such tetragonal-distortion centres is substantial, owing to their large associated shear strains⁵¹. Generally, hardening due to tetragonal distortion fields is much larger than that for spherical distortions (for example, substitutional solid-solution strengthening). Fleischer treated solid-solution hardening by such tetragonal distortion fields as ‘rapid hardening’ and estimated the yield strength increment to be:

$$\Delta\tau = \frac{G\Delta\epsilon c^{1/2}}{3} \quad (2)$$

where G is the shear modulus, $\Delta\epsilon$ is the difference between the longitudinal and transverse strain of the tetragonal distortion source when interstitial atoms occupy the interstitial sites in alloys, and c is the atomic concentration of interstitial atoms creating such defects. Herein we use the change of hardness ΔH_v to quantify the effect of solid-solution strengthening⁵², rendering equation (2) thus:

$$\Delta H_v = 3^{3/2} \frac{G\Delta\epsilon c^{1/2}}{3} = 3^{1/2} G\Delta\epsilon c^{1/2} \quad (3)$$

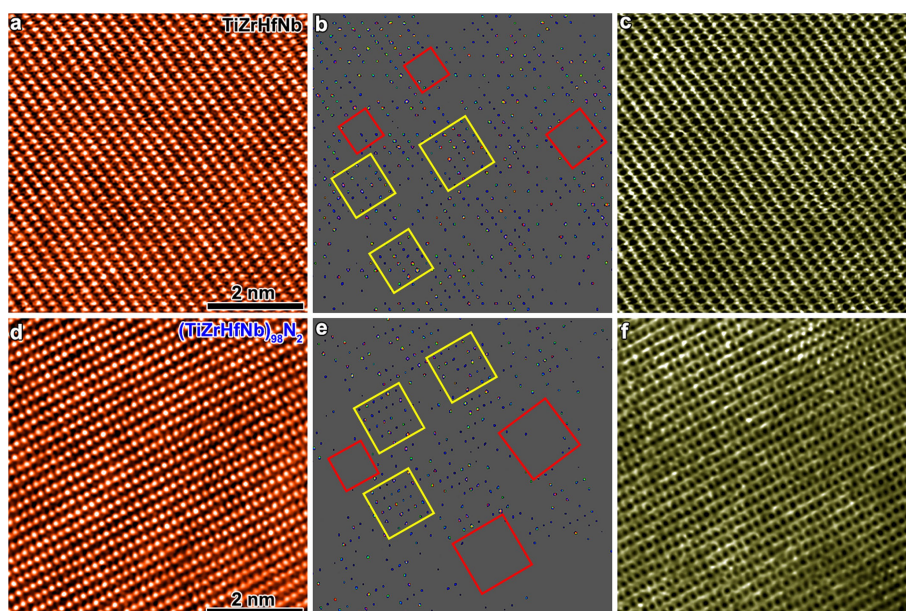
where 3^{3/2} is a conversion factor between shear stress and hardness. The Young’s modulus of the base HEA is 110.2 ± 7.9 GPa, obtained from nanoindentation measurements and the shear modulus was calculated to be 42.4 GPa. Based on these parameters, the calculated values of the tetragonal distortion $\Delta\epsilon$ for the O-2 and N-2 HEAs are 0.10 and 0.14, respectively. These values are comparable to those of other tetragonal lattice distortions at room temperature⁵³, indicating that the hardening mechanism in the current alloys is indeed of interstitial nature. The larger atomic size of nitrogen (atomic radius of 0.75 Å) compared to that of oxygen (0.65 Å) causes a higher asymmetry of the tetragonal distortions, thus leading to a larger $\Delta\epsilon$ value and hence more pronounced interstitial strengthening. Via extrapolation from the synchrotron XRD data, the lattice parameter of the base, O-2 and N-2 HEAs was estimated to be 3.4308 Å, 3.4331 Å and 3.4347 Å, respectively, which confirms that the lattice distortion associated with adding nitrogen is indeed larger than that introduced when adding oxygen (Extended Data Fig. 10). As expected by conventional interstitial strengthening theory^{54–62}, the interstitial strengthening always suffers from the strength–ductility trade-off (that the increase in strength leads to decreasing ductility). Nevertheless, addition of oxygen in the current b.c.c. HEA successfully reverses the strength–ductility trade-off.

Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

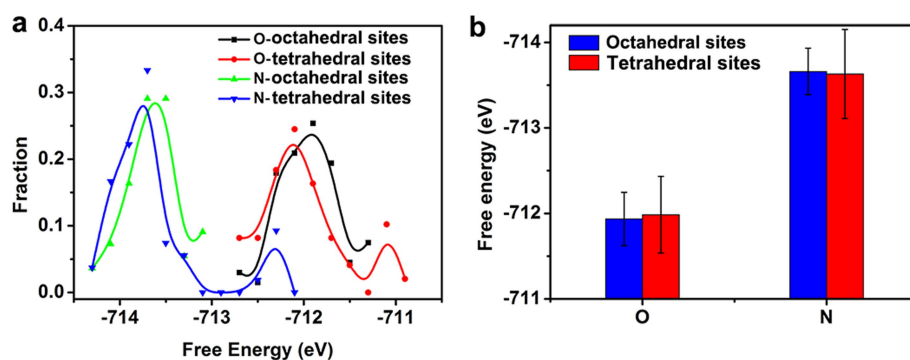
33. Granberg, F. et al. Mechanism of radiation damage reduction in equiatomic multicomponent single phase alloys. *Phys. Rev. Lett.* **116**, 135504 (2016).
34. Zhang, Z. et al. Nanoscale origins of the damage tolerance of the high-entropy alloy CrMnFeCoNi. *Nat. Commun.* **6**, 10143 (2015).

35. Zhang, Y. et al. Influence of chemical disorder on energy dissipation and defect evolution in concentrated solid solution alloys. *Nat. Commun.* **6**, 8736 (2015).
36. Zou, Y., Ma, H. & Spolenak, R. Ultrastrong ductile and stable high-entropy alloys at small scales. *Nat. Commun.* **6**, 7748 (2015).
37. Senkov, O. N. et al. Microstructure and elevated temperature properties of a refractory TaNbHfZrTi alloy. *J. Mater. Sci.* **47**, 4062–4074 (2012).
38. Gorr, B. et al. Phase equilibria, microstructure, and high temperature oxidation resistance of novel refractory high-entropy alloys. *J. Alloys Compd.* **624**, 270–278 (2015).
39. Wu, Y. D. et al. A refractory Hf₂₅Nb₂₅Ti₂₅Zr₂₅ high-entropy alloy with excellent structural stability and tensile properties. *Mater. Lett.* **130**, 277–280 (2014).
40. Thompson, K. et al. In situ site-specific specimen preparation for atom probe tomography. *Ultramicroscopy* **107**, 131–139 (2007).
41. An, K. *VDRIVE-Data Reduction and Interactive Visualization Software for Event Mode Neutron Diffraction*. ORNL Report No. ORNL-TM-2012-621 (Oak Ridge National Laboratory, Oak Ridge, 2012).
42. Grandini, C. R. A low cost automatic system for anelastic relaxations measurements. *Rev. Brasil. Apl. Vácuo* **21**, 13–16 (2008).
43. Nowick, A. S. *Anelastic Relaxation in Crystalline Solids* Vol. 1 (Elsevier, New York, 2012).
44. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
45. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).
46. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
47. Zunger, A., Wei, S. H., Ferreira, L. & Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* **65**, 353 (1990).
48. Press, W. H. *The Art of Scientific Computing* (Cambridge Univ. Press, New York, 1992).
49. Blöchl, P. E., Jepsen, O. & Andersen, O. K. Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B* **49**, 16223 (1994).
50. Methfessel, M. & Paxton, A. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B* **40**, 3616 (1989).
51. Courtney, T. H. *Mechanical Behaviour of Materials* (Waveland Press, New York, 2005).
52. Schuh, C. A., Nieh, T. G. & Iwasaki, H. The effect of solid solution W additions on the mechanical properties of nanocrystalline Ni. *Acta Mater.* **51**, 431–443 (2003).
53. Mitchell, T. E. & Heuer, A. H. Solution hardening by aliovalent cations in ionic crystals. *Mater. Sci. Eng. A* **28**, 81–97 (1977).
54. Finlay, W. L. & Snyder, J. A. Effects of three interstitial solutes (nitrogen, oxygen, and carbon) on the mechanical properties of high-purity, alpha titanium. *J. Met.* **2**, 277–286 (1950).
55. Ulitschny, M. & Gibala, R. The effects of interstitial solute additions on the mechanical properties of niobium and tantalum single crystals. *J. Less Common Met.* **33**, 105–116 (1973).
56. Šob, M., Kratochvíl, J. & Kroupa, F. Theory of strengthening of alpha titanium by interstitial solutes. *Czech. J. Phys.* **25**, 872–890 (1975).
57. Nakada, Y. & Keh, A. S. Solid-solution strengthening in Ni-C alloys. *Metall. Trans.* **2**, 441–447 (1971).
58. Nakada, Y. & Keh, A. S. Solid solution strengthening in Fe-N single crystals. *Acta Metall.* **16**, 903–914 (1968).
59. Li, Y. J., Ponge, D., Choi, P. & Raabe, D. Segregation of boron at prior austenite grain boundaries in a quenched martensitic steel studied by atom probe tomography. *Scr. Mater.* **96**, 13–16 (2015).
60. Kim, M., Geller, C. B. & Freeman, A. J. The effect of interstitial N on grain boundary cohesive strength in Fe. *Scr. Mater.* **50**, 1341–1343 (2004).
61. San Marchi, C. et al. in *Proc. 2008 International Hydrogen Conf.* 88–96 (ASM International, Russell Township, 2008).
62. Dadfarnia, M. et al. Recent advances in the study of structural materials compatibility with hydrogen. *Adv. Mater.* **22**, 1128–1135 (2010).
63. Snoek, J. Effect of small quantities of carbon and nitrogen on the elastic and plastic properties of iron. *Physica* **8**, 711–733 (1941).
64. Nelson, J. B. & Riley, D. An experimental investigation of extrapolation methods in the derivation of accurate unit-cell dimensions of crystals. *Proc. Phys. Soc.* **57**, 160 (1945).



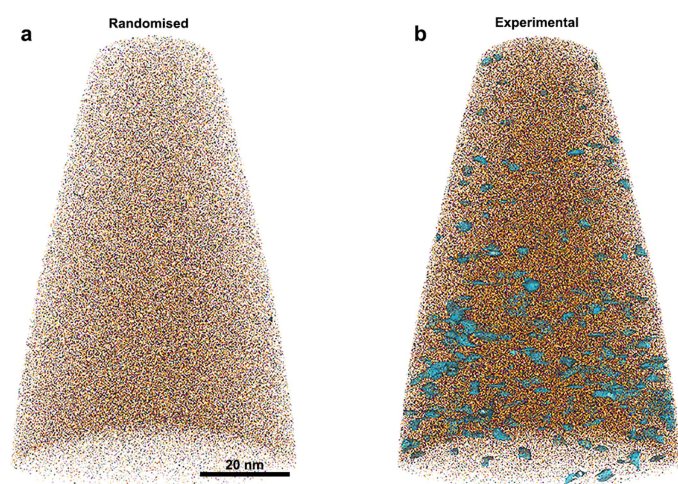
Extended Data Fig. 1 | Aberration-corrected STEM of the as-cast HEAs. Shown are the HAADF-STEM images for the $[011]_{\text{b.c.c.}}$ crystal axis with differently adjusted contrasts to show the existence of chemical short-range ordering, and the corresponding STEM-ABF images, for the equiatomic TiZrHfNb high-entropy base alloy (**a–c**) and for N-2 HEA

(that is, $(\text{TiZrHfNb})_{98}\text{N}_2$) (**d–f**). The red panel represents the Zr/Ti-rich region, while the yellow panel indicates the Hf/Nb-rich region. No ordered interstitial occupation is observed in these two HEAs. Red squares represent the Zr/Ti-rich region and yellow squares indicate the Hf/Nb-rich region.

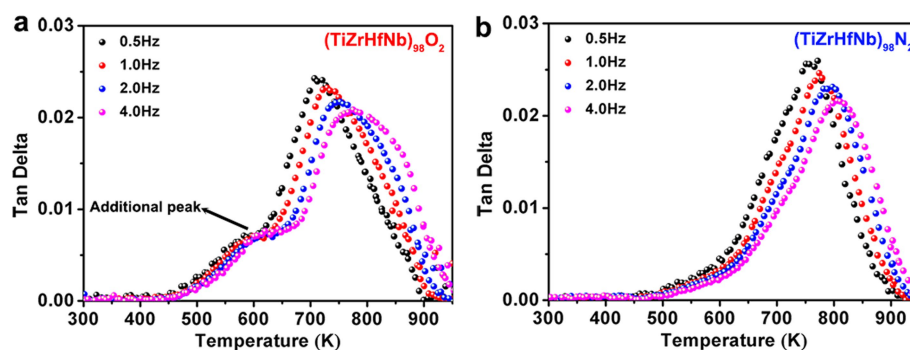


Extended Data Fig. 2 | Occupation possibility analysis of interstitial oxygen/nitrogen from first-principles calculations. a, Statistical distribution of system energy for the case of oxygen/nitrogen at different interstitial sites in the TiZrHfNb HEA. **b,** Comparison of average free energy for the systems with oxygen/nitrogen atoms at octahedral and tetrahedral interstitial sites. It can be seen that the octahedral interstitial

oxygen/nitrogen has a free energy nearly identical to that of the tetrahedral interstitial oxygen/nitrogen, indicating that the likelihood of oxygen/nitrogen atoms occupying the tetrahedral or octahedral interstitial sites in the b.c.c. lattice is similar. The error bars represent the standard error of the mean.

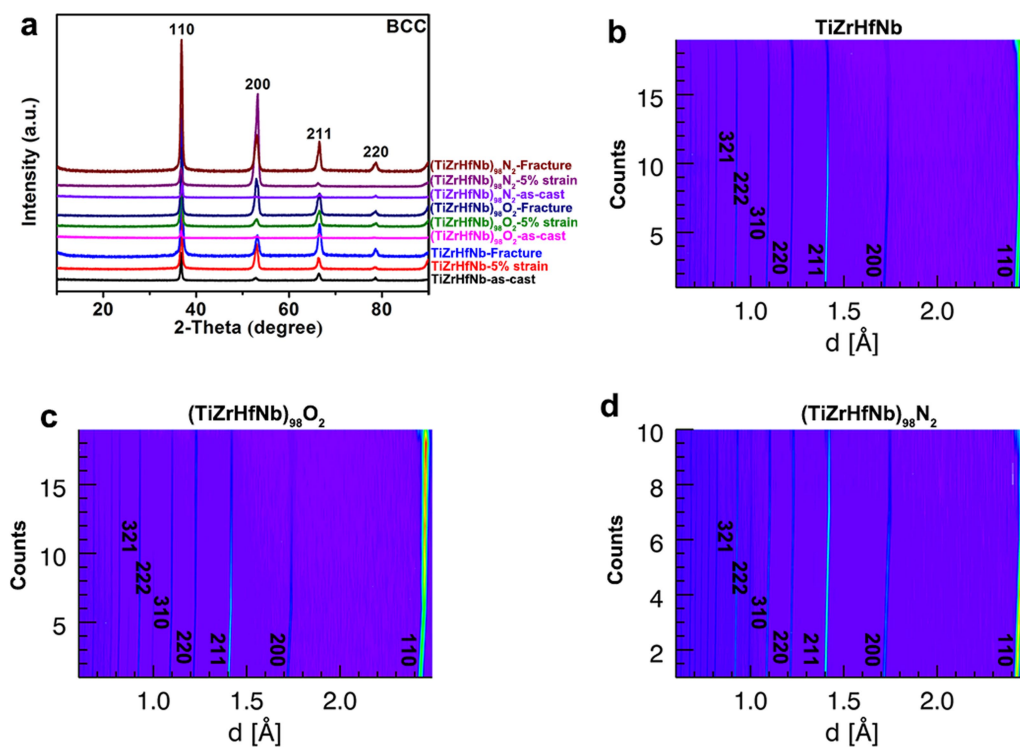


Extended Data Fig. 3 | Three-dimensional reconstruction of the O-2 HEA atom probe tomography dataset. a, b, Randomized (a) and experimental (b) datasets on which an iso-composition surface encompassing regions in the point cloud containing more than 3.0 at% O was superimposed. The experimental dataset clearly shows evidence for OOCs.



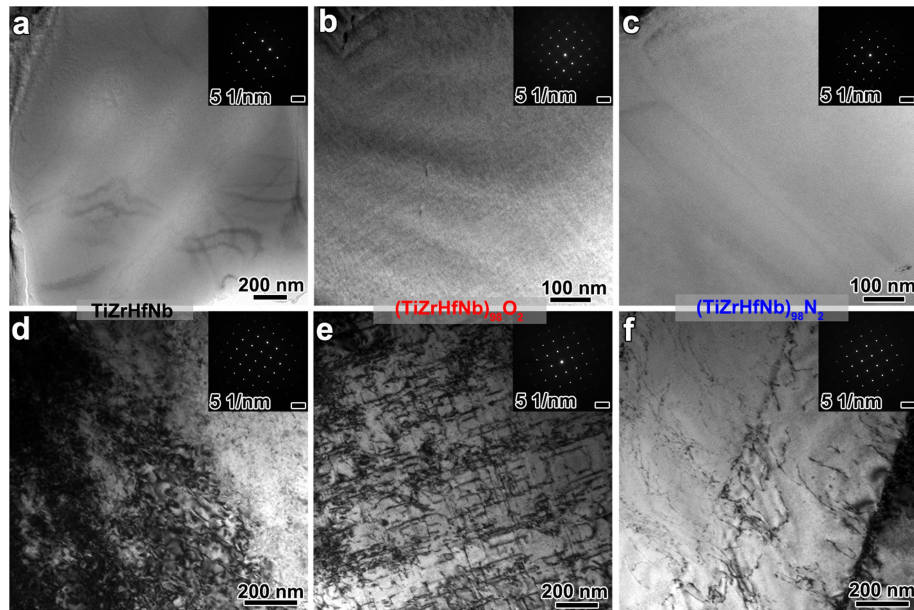
Extended Data Fig. 4 | Internal-friction measurements. Internal-friction results obtained for the O-2 (that is, $(\text{TiZrHfNb})_{98}\text{O}_2$) and N-2 (that is, $(\text{TiZrHfNb})_{98}\text{N}_2$) HEAs. Metals containing solute atoms in interstitial solution show Snoek relaxation behaviour owing to stress-induced ordering⁴³, which gives rise to a peak in the corresponding internal-friction spectrum (that is, the Snoek peak)⁶³, and different internal-friction peaks correspond to different types of local atomic environments

of the interstitials. Tan Delta represents the damping capacity. **a**, For the oxygen-doped alloy $(\text{TiZrHfNb})_{98}\text{O}_2$ two peaks are observed: a dominant high-temperature peak and an additional low-temperature peak. **b**, For the nitrogen-doped alloy $(\text{TiZrHfNb})_{98}\text{N}_2$ only the main peak is observed. This observation suggests that addition of oxygen to the TiZrHfNb HEA induces formation of two different types of interstitial atomic structures, unlike in the N-2 HEA, where only a single solid-solution peak appears.



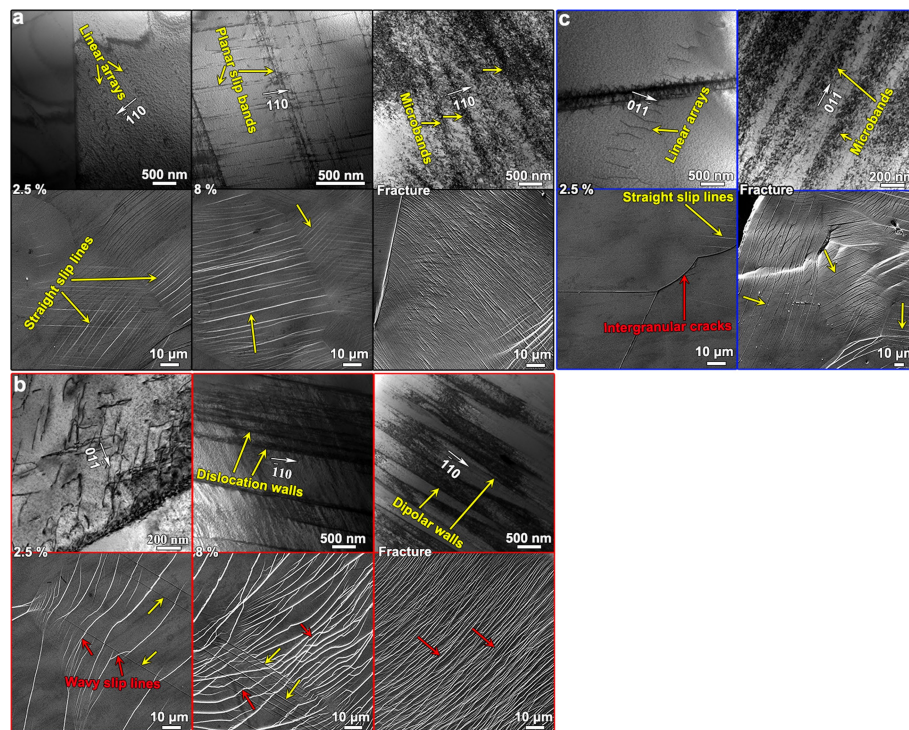
Extended Data Fig. 5 | X-ray and in situ neutron diffraction measurements. **a**, XRD patterns of the TiZrHfNb base alloy, O-2 HEA (that is, $(\text{TiZrHfNb})_{98}\text{O}_2$) and N-2 HEA (that is, $(\text{TiZrHfNb})_{98}\text{N}_2$) with different pre-tensioned strains. **b–d**, In situ neutron diffraction patterns

of the three alloys. d is the interplanar distance. Both ex situ XRD and in situ neutron diffraction measurements confirm that there is no phase transformation in the three HEAs during deformation.



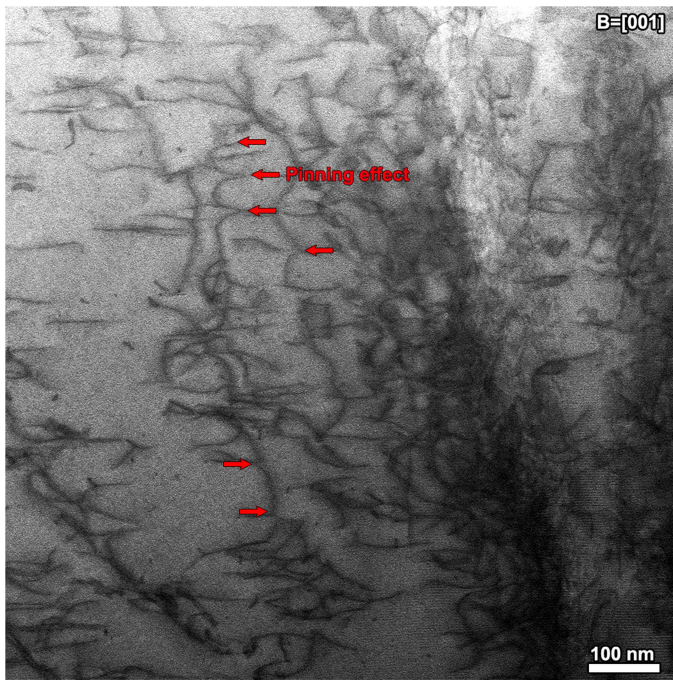
Extended Data Fig. 6 | Transmission electron microscopy. **a–c**, TEM images of the as-cast equiatomic TiZrHfNb base alloy, O-2 HEA (that is, $(\text{TiZrHfNb})_{98}\text{O}_2$) and N-2 HEA (that is, $(\text{TiZrHfNb})_{98}\text{N}_2$). **d–f**, TEM images of the fractured HEA specimens. The TEM results further confirm

that no second phase appears before and after the tensile tests. The inset in each figure is the corresponding electron diffraction pattern of the selected area.

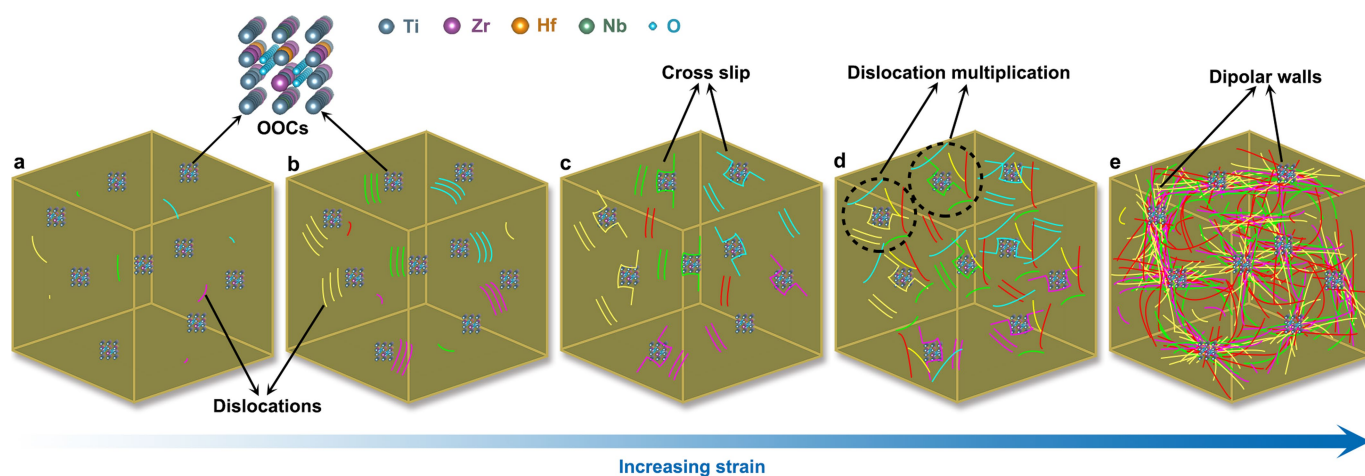


Extended Data Fig. 7 | Dislocation configuration. **a**, For the equiatomic TiZrHfNb base alloy, at low tensile strain (2.5% strain), dislocations in linear arrays are observed. As the strain increases to 8%, planar slip bands and individual dislocation-rich sheets are formed. After fracture, although irregular dislocation cells can be seen, there exist several microbands, indicating that planar slip is still the dominant deformation mode. **b**, For the oxygen-doped alloy variant O-2 HEA (that is, (TiZrHfNb)₉₈O₂) at 2.5% strain, however, the dislocations are arranged in bundles and loops. At 8% strain, dislocation walls are formed. For the dislocation substructure after fracture, dipolar walls that mainly contain primary dislocation dipoles at a high density are observed, suggesting a typical cell-forming

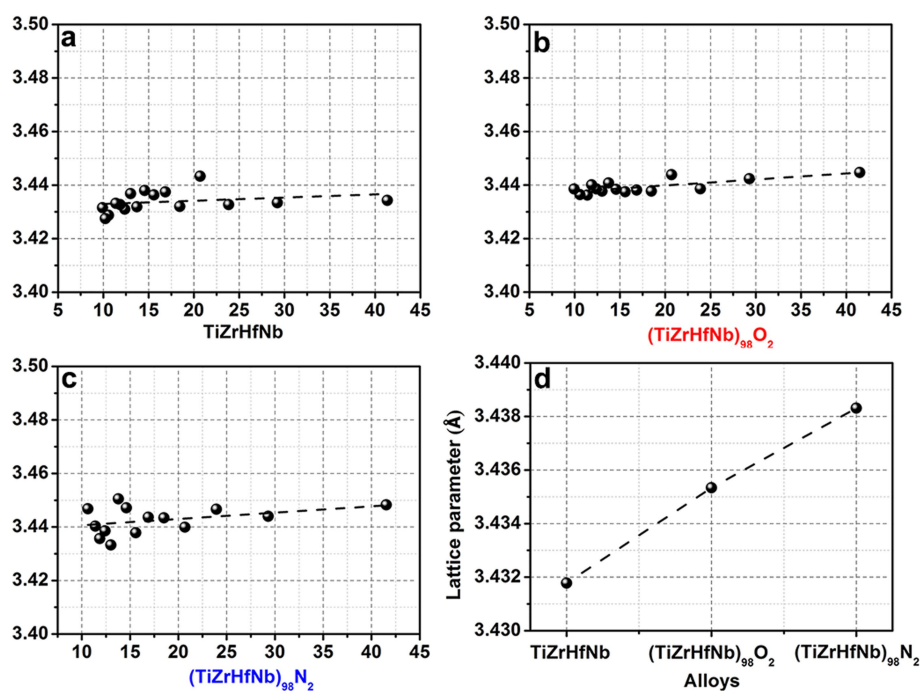
deformation microstructure in the O-2 HEA. **c**, For the nitrogen-doped alloy N-2 HEA (that is, (TiZrHfNb)₉₈N₂), the deformation mode is similar to that of the base alloy. In addition, slip traces at each specimen surface during deformation are also observed. Even at a low strain (2.5% strain), wavy slip lines are clearly seen in the oxygen doped variant O-2 HEA, whereas in the TiZrHfNb base alloy, even at high strain (8%), straight slip lines prevail and wavy slip lines only occur upon necking. Moreover, premature and much more serious necking occurs in the TiZrHfNb base alloy. It is worth mentioning here that intergranular fracture is observed in the nitrogen-doped variant N-2 HEA, which is probably caused by grain boundary segregation of nitrogen.



Extended Data Fig. 8 | Pinning effect. Aberration-corrected STEM observation of O-2 HEA (that is, (TiZrHfNb)₉₈O₂) after being pre-strained to 8%. *B* is the beam direction. The red arrows indicate the distinct dislocation pinning effect, which suppresses dislocation motion substantially during deformation.



Extended Data Fig. 9 | Schematic diagram illustrating the plastic deformation mechanism in the oxygen-rich alloy variant O-2 HEA.



Extended Data Fig. 10 | Lattice parameter calculation. a–c, Plot of measured values of the lattice parameter versus $\frac{\cos^2 \vartheta}{2} \left(\frac{1}{\sin \vartheta} + \frac{1}{\vartheta} \right)$ for TiZrHfNb, (TiZrHfNb)₉₈O₂ and (TiZrHfNb)₉₈N₂ alloys. The position of each peak was measured on the diffractogram from which the lattice parameter was calculated. The measured lattice parameters were plotted

versus $\frac{\cos^2 \vartheta}{2} \left(\frac{1}{\sin \vartheta} + \frac{1}{\vartheta} \right)$, where ϑ is the Bragg angle for each peak and the resulting graph extrapolated to zero to obtain the best value of the lattice parameter⁶⁴. d, The calculated lattice parameters of the three alloys.

Twentieth-century contribution to sea-level rise from uncharted glaciers

David Parkes^{1,2*} & Ben Marzeion²

Global-mean sea-level rise (GMSLR) during the twentieth century was primarily caused by glacier and ice-sheet mass loss, thermal expansion of ocean water and changes in terrestrial water storage¹. Whether based on observations² or results of climate models^{3,4}, however, the sum of estimates of each of these contributions tends to fall short of the observed GMSLR. Current estimates of the glacier contribution to GMSLR rely on the analysis of glacier inventory data, which are known to undersample the smallest glacier size classes^{5,6}. Here we show that from 1901 to 2015, missing and disappeared glaciers produced a sea-level equivalent (SLE) of approximately 16.7 to 48.0 millimetres. Missing glaciers are those small glaciers that we expect to exist today, owing to regional analyses and theoretical scaling relationships, but that are not represented in the inventories. These glaciers contributed approximately 12.3 to 42.7 millimetres to the historical SLE. Additionally, disappeared glaciers (those that existed in 1901 but had melted away by 2015, and that therefore cannot be included in modern global glacier inventories) made an estimated contribution of between 4.4 and 5.3 millimetres. Failure to consider these uncharted glaciers may be an important cause of difficulties in closing the GMSLR budget during the twentieth century: their contribution is on average between 0.17 and 0.53 millimetres of SLE per year, compared to a budget discrepancy of about 0.5 millimetres of GMSLR per year between 1901 and 1990. Although the uncharted glaciers will have a minimal role in sea-level rise in the future, and are less important after 1990, these findings imply that undiscovered physical processes are not required to close the historical sea-level budget.

Mass loss from glaciers forms a major component of GMSLR during the twentieth century¹. Direct historical records of glacier mass changes are small in number compared to the total number of glaciers globally^{7,8}, so methods for upscaling these observations to the global scale are necessary for assessing the GMSLR budget^{2–4}. The available methods cover a wide range of complexity, from geographically weighted interpolation⁸, to scaling from glacier length change observations^{9,10}, to numerical modelling of each individual glacier based on climate observations¹¹. All these methods rely on comprehensive global inventories of glaciers, which are a relatively recent development made possible by large-scale aerial mapping¹² and satellite-based Earth observation techniques⁶. Glacier inventories are therefore only reasonably representative of current or recent glacier states, with not enough information available to determine historical states. Accuracy in reconstructed SLE mass change contribution from glaciers is limited by the effective ‘resolution’ of the glacier inventory (the minimal glacier size the inventory can faithfully represent), which underpins the reconstructive method. Accuracy is further limited by the possibility that glaciers that have already completely disappeared contributed to mass change in the past. There is strong evidence that small glaciers are also under-represented in the most up-to-date inventories, compared to expected glacier distributions^{5,6,13}. In some regions, glaciers sized below a certain threshold are deliberately excluded from the inventories¹⁴. Improvements in remote observation techniques alone are not

an efficient way to reduce the limitation that glacier inventory resolution places on global glacier reconstructions: new and improved datasets are expensive, time-consuming to collect (requiring lots of manual labour)¹⁵, and limited by available sensing technologies and the missions that use them. Furthermore, reducing the error in global or regional total glacier mass by an order of magnitude could require improving the effective resolution of glacier inventories by almost four orders of magnitude⁵, which would necessitate a huge advance in remote sensing. It is important to note that any error in the present-day representation of small glaciers results in proportionally larger errors in reconstructions of these glaciers’ change during the past, because small glaciers tend to have experienced much greater proportional changes in volume and area since pre-industrial times than have larger glaciers (Extended Data Fig. 1b). Methods to account for the limitations of glacier inventories without explicitly collecting the missing data are therefore of great interest for improving global glacier mass-change estimates, and as we suggest here, for closing the GMSLR budget.

First, we define two new classes of glaciers that need to be considered when estimating the glaciers’ SLE mass change. ‘Missing’ glaciers are those that we expect to exist in 2015, but which are not contained in the Randolph Glacier Inventory (RGI) version 5 (RGIv5)¹⁶, the 2015 release of the RGI, with data for over 200,000 glaciers: this under-sampling is due to limitations in remote sensing methods. ‘Disappeared’ glaciers are those that we expect to have existed in 1901, but that melted entirely away between 1901 and 2015: they are a contribution systematically left out by glacier reconstructions that rely explicitly on modern inventories, regardless of the quality of remote sensing data. We use the term ‘uncharted’ glaciers to refer to the combination of these two classes of glacier.

We then combine glacier modelling and empirically determined global power laws relating glacier frequency density and glacier surface area S to estimate the 1901–2015 SLE mass loss contribution for missing and disappeared glaciers. The existence of a power-law relationship between glacier surface area and frequency density is supported by theoretical evidence¹³ and observational evidence on a regional scale⁵, as well as evidence of a similar power law holding for smaller snow-deposition-based phenomena such as snow patches¹³. We find strong evidence for the same form of power law holding globally (Fig. 1). From RGIv5 data, the power law holds globally for glaciers between $10^{0.3}$ (about 2) km^2 and $10^{2.6}$ (about 398) km^2 , with the fall-off in frequency density for large glaciers a consequence of the limitations in size and topography of glacierized regions. The fall-off in frequency density for small glaciers ($10^{-2.0}$ ($=0.01$) $\text{km}^2 \leq S < 10^{0.3}$ km^2 , where the lower limit of $10^{-2.0}$ km^2 is the minimum glacier size in RGIv5), however, does not have a known physical justification. As the power law holds both for a wide range of mid-sized glaciers, and also for smaller but similarly distributed phenomena such as snow patches⁵, and since there is no posited mechanism reducing the occurrence of small glaciers, the fall-off at small glacier sizes has been hypothesized to be explained by under-representation in the global inventory^{5,6}. On the basis of this hypothesis, we derive an upper-bound estimate of the contribution of uncharted glaciers.

¹Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria. ²Institute of Geography, University of Bremen, Bremen, Germany.

*e-mail: david.parkes.88@gmail.com

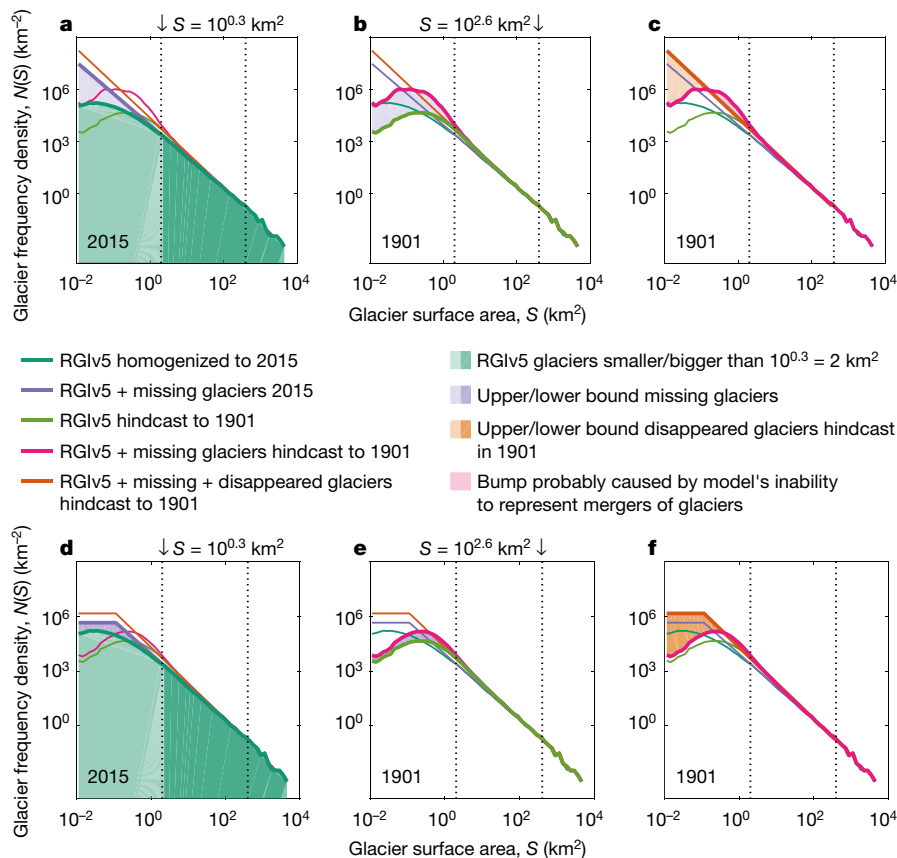


Fig. 1 | Frequency density of glaciers as a function of glacier size. The top and bottom rows show the upper and lower bounds respectively. Within each row, each panel shows the same set of distributions, but highlights different distributions, representative of a different set of glaciers. **a, d**, The RGIv5 glacier distribution in 2015 (split into light/dark

green for small/larger glaciers) and the power-law-derived distribution of missing glaciers in 2015. **b, e**, The distribution of missing glaciers in 1901. **c, f**, The 1901 distribution of disappeared glaciers, as well as a ‘bump’ that we consider to be a modelling artefact.

The hypothesis that the power law holds down to the smallest glacier size classes remains to be tested, so we also derive a lower-bound estimate of the contribution of uncharted glaciers. The lower bound assumes that there is a cut-off in glacier size for which the power law holds, and that the frequency density of glaciers smaller than the cut-off, as a function of area, is constant. The chosen cut-off of 0.1 km^2 is an order of magnitude larger than the minimum glacier size recorded in RGIv5, implying that the lower-bound estimate is considerably less affected by potential effects from ice bodies at the glacier or snow-patch transition. These borderline glaciers are not necessarily dominated by the same physical processes as larger glaciers, so scaling relationships may be less certain. The upper and lower bounds account for a range of possibilities of partial flattening or tailing-off of the power law at the smallest glacier sizes. Observations in Switzerland show power-law behaviour extending down to smaller glacier sizes than in the global dataset, but with a smaller exponent (Methods and Extended Data Fig. 2). If the Swiss data are globally representative, this suggests a reality falling comfortably between the upper and lower bounds.

To account for missing glaciers, we scale up the mass change from small glacier size classes in RGIv5. For the upper bound, using the power law observed for glaciers between $10^{0.3} \text{ km}^2$ and $10^{2.6} \text{ km}^2$, we generate an upscaling factor for each glacier size class below $10^{0.3} \text{ km}^2$ equal to the ratio of the power-law-predicted 2015 frequency density to the 2015 RGIv5 frequency density. The lower bound has size classes below 0.1 km^2 scaled to the frequency density at 0.1 km^2 . By hindcasting the annual mass change for each glacier in RGIv5 back to 1901 using an established global glacier model¹¹ that has been used extensively in sea-level budget assessment^{2–4,17,18}, and applying these upscaling factors to the contribution of each small glacier (‘small’ according to their 2015 size), we generate mass change estimates for missing glaciers

between 1901 and 2015. As the 1901 glaciers are also expected to be distributed following a power law, we independently fit a power law to the RGIv5 + missing glaciers in 1901, and this time upscale the 1901 mass, using newly generated upscaling factors, to account for the total mass of disappeared glaciers. The glacier mass added by this second upscaling is expected to have entirely disappeared by 2015, so by combining the total mass of disappeared glaciers and the 1901–2015 mass change from missing glaciers, we arrive at upper- and lower-bound total mass change estimates for all uncharted glaciers during the period 1901–2015.

The RGIv5 glacier frequency distribution for 2015 (Fig. 1a, d, dark green line) gives a power-law exponent of -1.80 ± 0.01 (Fig. 1a, d, purple line), resulting in an upper/lower bound of $42.7 \pm 6.5/12.3 \pm 1.6 \text{ mm SLE}$ (95% confidence interval; see the ‘Estimation of errors’ section in the Methods for details) mass loss between 1901 and 2015 from missing glaciers. The power-law fit applied to the 1901 glacier distribution including upper-bound missing glaciers (Fig. 1b, pink line) gives an exponent of -1.98 ± 0.04 (Fig. 1c, orange line), resulting in an expected upper bound of $5.3 \pm 2.4 \text{ mm SLE}$ mass loss over the same period from disappeared glaciers. Applied to the lower 1901 glacier distribution with lower-bound missing glaciers (Fig. 1e, pink line) the power-law fit gives an exponent of -1.96 ± 0.03 (Fig. 1f, orange line), and a lower bound of $4.4 \pm 1.4 \text{ mm SLE}$ mass loss from disappeared glaciers. We note that the different exponents for 2015 and 1901 may be because of the state of glaciers relative to equilibrium, as the response of smaller glaciers is faster than the response of large glaciers, resulting in a ‘flattening’ of the distribution as glaciers in general shrink. Bahr and Radic⁵ find a regionally averaged (across ten glacierized regions) exponent of -2.10 ± 0.09 , and the theoretical exponent given by Bahr and Meier¹³ is -2.05 , implicitly for an equilibrium scenario. We see a

Table 1 | Breakdown of current ice mass and mass changes

	Total ice mass in 2015 (mm SLE)	Mass loss contribution 1901–2015 (mm SLE)
RGIv5 glaciers $\geq 10^{0.3}$ km ²	489.3 \pm 21.4	75.1 \pm 3.3
RGIv5 glaciers $< 10^{0.3}$ km ²	3.2 \pm 0.1	14.0 \pm 0.6
RGIv5 total	492.5 \pm 21.6	89.1 \pm 3.9
Missing glaciers (upper bound)	2.4 \pm 0.4	42.7 \pm 6.5
Disappeared glaciers (upper bound)	0	5.3 \pm 2.4
Uncharted glaciers (upper bound)	2.4 \pm 0.4	48 \pm 8.9
Missing glaciers (lower bound)	2.1 \pm 0.3	12.3 \pm 1.6
Disappeared glaciers (lower bound)	0	4.4 \pm 1.4
Uncharted glaciers (lower bound)	2.1 \pm 0.3	16.7 \pm 3.0

Current ice masses and 1901–2015 SLE mass loss contributions from different glacier subsets (95% uncertainty ranges). Boldface rows are the sum of the previous two rows. In the RGIv5 results—as in the missing glaciers—we see that a small glacier mass in 2015 was responsible for a much larger proportion of historical glacier mass loss than their 2015 mass may suggest, as these glaciers have typically seen a much greater proportionate mass change than large glaciers (see also Extended Data Fig. 1b). Disappeared glaciers, by definition, do not exist in 2015, but still contributed a modest amount to SLE mass loss.

slightly flatter distribution in 1901 and a flatter distribution still in 2015, qualitatively in agreement with general glacier shrinkage during the twentieth century.

Combining the contributions of missing and disappeared glaciers, we derive a total mass loss upper bound of 48.0 ± 8.9 and a lower bound of 16.7 ± 3.0 mm SLE from uncharted glaciers. The revised mass change figure from the reconstruction of all RGIv5 glaciers using the unmodified glacier model¹¹ forced with the climate observations¹⁹ (Climate Research Unit (CRU) version 3.24) and initialized using RGIv5¹⁶ is 89.1 ± 3.9 mm SLE. Uncharted glaciers have therefore contributed 35.0% of a total 137.1 ± 12.8 mm SLE glacier mass loss between 1901 and 2015 (using upper bound values), and 15.8% of a total 105.8 ± 6.9 mm SLE (using lower bound values). The SLE 2015 mass and 1901–2015 mass loss contribution for each class of glacier is summarized in Table 1.

The uncharted glacier contribution to GMSLR over the period 1901–2015 is estimated to be between 0.15 and 0.42 mm of SLE per year, and the total glacier contribution during the same period to be between 0.93 and 1.20 mm of SLE per year. The upper bound uncharted contribution may close the sea-level budget discrepancy identified in the IPCC's Fifth Assessment Report³ for 1901 to 1990 (0.17/0.53 mm of SLE per year lower/upper bound contribution compared to a discrepancy of 0.5 mm of GMSLR per year), but only covers part of the discrepancy for 1993 to 2010 (0.08/0.21 mm of SLE per year lower/upper bound contribution compared to a discrepancy of 0.4 mm of GMSLR per year). Smaller estimates of twentieth-century GMSLR have been published since the IPCC Fifth Assessment Report^{20,21}. However, because their methods (to different degrees) depend on the sea-level fingerprint of glacier mass loss, the impact of our results on a sea-level budget closure based on these recent GMSLR estimates is not immediately obvious.

The modelled annual SLE mass loss contribution between 1901 and 2015 for RGIv5, missing, and disappeared glaciers is shown in Fig. 2. The data from RGIv5 glaciers is split into the contribution from small glaciers (which contributes to the upscaling of the uncharted glaciers) and that from larger glaciers. This shows that the contribution of uncharted glaciers, both in absolute terms and as a proportion of total glacier contribution, is largest early in the twentieth century. The contribution decreases gradually to the point where it is negligible by 2015, with the total remaining volume of missing glaciers in 2015 comprising only between 2.1 ± 0.3 and 2.4 ± 0.4 mm of SLE ice mass. This potential contribution is likely to be realized in the very near future, as missing glaciers are very small, and from past surface mass balances (Extended Data Fig. 1a) we can see that small glaciers tend to have more rapid proportionate mass changes. The decreasing pattern of mass loss for uncharted glaciers is largely distinct from the overall pattern of RGIv5 mass change, which shows increasing mass loss rates up to a peak in the 1930s, decreasing to a minimum around 1970, and subsequent increase until the 2010s (this temporal pattern is the result of spatially inhomogeneous climate variability during the twentieth century¹¹).

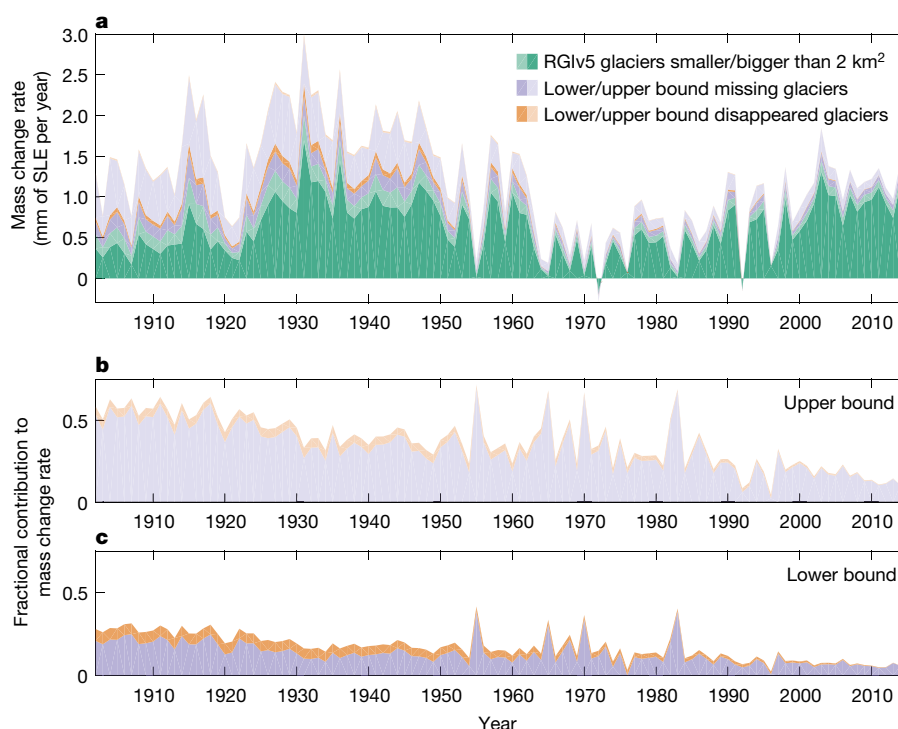


Fig. 2 | Annual glacier mass loss time series. The light and dark green sections show the hindcast mass loss from RGIv5 glaciers. The additional contribution from missing glaciers is shown in purple, and that of disappeared glaciers is shown in orange. **a**, The lower and upper bounds are stacked (light and dark purple combined give the upper bound of the

missing glaciers' contribution, light and dark orange combined that of the disappeared glaciers). **b**, **c**, The missing and disappeared glaciers' upper-bound (**b**) and lower-bound (**c**) contributions, respectively, are separated, and the fractional mass changes are calculated based on the separate totals.

The consideration of uncharted glaciers adds a substantial amount to hindcast SLE mass loss contributions from glaciers between 1901 and 2015 that is comparable to the existing discrepancy between known GMSLR contributors and observed sea-level change. It is therefore imperative for the closure of the twentieth-century GMSLR budget that these glaciers are considered. Although higher-quality observational datasets can reduce the need for upscaling to account for uncharted glaciers, these are limited by the timeframes and technical development of Earth observation missions. In the case of disappeared glaciers, even a theoretically perfect inventory still cannot represent their mass changes. Accounting for uncharted glaciers thus cannot be done exclusively through improvements in glacier inventories, and upscaling (or other methods) of contributions from known glaciers to account for glaciers outside of either the resolution or scope of current glacier inventories must form an integral part of accurate GMSLR hindcasting.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0687-9>.

Received: 1 September 2017; Accepted: 10 September 2018;

Published online 21 November 2018.

- Church, J. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. et al.) Ch. 13 (IPCC, Cambridge Univ. Press, Cambridge/New York, 2013).
- Gregory, J. M. et al. Twentieth-century global-mean sea-level rise: is the whole greater than the sum of the parts? *J. Clim.* **26**, 4476–4499 (2013).
- Church, J. A., Monselesan, D., Gregory, J. M. & Marzeion, B. Evaluating the ability of process based models to project sea-level change. *Environ. Res. Lett.* **8**, 014051 (2013).
- Slangen, A. et al. Anthropogenic forcing dominates global mean sea-level rise since 1970. *Nat. Clim. Change* **6**, 701–705 (2016).
- Bahr, D. & Radic, V. Significant contribution to total mass from very small glaciers. *Cryosphere* **6**, 763–770 (2012).
- Pfeffer, W. T. et al. The Randolph Glacier Inventory: a globally complete inventory of glaciers. *J. Glaciol.* **60**, 537–552 (2014).
- Zemp, M. et al. Historically unprecedented global glacier decline in the early 21st century. *J. Glaciol.* **61**, 745–762 (2015).
- Cogley, J. G. Geodetic and direct mass-balance measurements: comparison and joint analysis. *Ann. Glaciol.* **50**, 96–100 (2009).
- Oerlemans, J., Dyurgerov, M. & van de Wal, R. S. W. Reconstructing the glacier contribution to sea-level rise back to 1850. *Cryosphere* **1**, 59–65 (2007).
- Leclercq, P. W., Oerlemans, J. & Cogley, J. G. Estimating the glacier contribution to sea-level rise for the period 1800–2005. *Surv. Geophys.* **32**, 519–535 (2011).
- Marzeion, B., Jarosch, A. H. & Hofer, M. Past and future sea-level change from the surface mass balance of glaciers. *Cryosphere* **6**, 1295–1322 (2012).
- WGMS (World Glacier Monitoring Service) and National Snow and Ice Data Center (NSIDC). *World Glacier Inventory* <http://nsidc.org/data/glacierinventory/index.html> (WGMS and NSIDC, 1989).
- Bahr, D. B. & Meier, M. F. Snow patch and glacier size distributions. *Water Resour. Res.* **36**, 495–501 (2000).
- Rastner, P. et al. The first complete inventory of the local glaciers and ice caps on Greenland. *Cryosphere* **6**, 1483–1495 (2012).
- Paul, F. et al. The glaciers climate change initiative: methods for creating glacier area, elevation change and velocity products. *Remote Sens. Environ.* **162**, 408–426 (2015).
- Arendt, A. et al. *Randolph Glacier Inventory—a Dataset of Global Glacier Outlines: Version 5.0* <https://www.glims.org/RGI/randolph50.html> (Global Land Ice Measurements from Space, Boulder, 2015).
- Frederikse, T. et al. Closing the sea level budget on a regional scale: trends and variability on the Northwestern European continental shelf. *Geophys. Res. Lett.* **43**, 10864–10872 (2016).
- Marcos, M. et al. Internal variability versus anthropogenic forcing on sea level and its components. *Surv. Geophys.* **38**, 329–348 (2017).
- Harris, I., Jones, P., Osborn, T. & Lister, D. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
- Hay, C. C., Morrow, E., Kopp, R. E. & Mitrovica, J. X. Probabilistic reanalysis of twentieth-century sea-level rise. *Nature* **517**, 481–484 (2015).
- Dangendorf, S. et al. Reassessment of 20th century global mean sea level rise. *Proc. Natl Acad. Sci. USA* **114**, 5946–5951 (2017).

Acknowledgements This research is funded by the Austrian Science Fund (FWF) project P25362.

Reviewer information Nature thanks W. Pfeffer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.P. and B.M. conceived and designed the study. B.M. performed the glacier model experiments. D.P. then developed and applied the upscaling techniques and performed the analysis. D.P. wrote the manuscript with contributions by B.M.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0687-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0687-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to D.P.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

The mass balance model used for hindcasting glacier evolution is the same as that described in ref. ¹¹, with the output updated for the RGIv5¹⁶ and CRU 3.24 climate observations¹⁹. The only modification is a change to the handling of data gaps to make it consistent with the treatment of uncharted glaciers: instead of assuming regional mean rates of glacier volume and area change for glaciers, which cannot be explicitly modelled (22.7% of all glacier area globally, of which 76.8% comes from Antarctic and Subantarctic peripheral glaciers; see Extended Data Table 1 for breakdown), the regional mean rates for glaciers within the same size class (as defined below) are assumed. For Antarctic and Subantarctic peripheral glaciers, where CRU data are not available so no glaciers can be explicitly modelled, global mean rates for glaciers within the same size class are assumed. See Methods section 'Impact of Antarctic peripheral glaciers' for a detailed consideration of the validity of using Antarctic peripheral glaciers in this study. We note that all SLE mass change values assume a global sea surface area of $3.619 \times 10^{14} \text{ m}^2$.

A 'base' run of the glacier model is established first. We require a snapshot of global glacier area distribution for a single point in time, but the observation years for RGIv5 glaciers differ. We model each glacier forward to 2015 if it has an observation year before 2015, to homogenize the data to a single snapshot in 2015 (distributed as shown in the dark green line in Fig. 1a, d), as well as hindcasting their sizes in 1901 (distributed as shown by the light green line in Fig. 1b, e).

The results for all glaciers are separated into size classes based on their modelled areas in 1901 and 2015. Size classes are logarithmic, defined as each set of glaciers with surface area S such that $10^{0.1i} \text{ km}^2 \leq S < 10^{0.1(i+1)} \text{ km}^2$ for each integer i with $-20 \leq i < 40$ (resulting in size classes that span from 0.01 km^2 to $10,000 \text{ km}^2$). As these size classes do not cover equal area ranges, we divide the glacier count in each size class (unitless) by the width of each size class (upper area limit minus lower area limit, in units of km^2) to get the glacier frequency density, $N(S)$, (in units of km^{-2}). It is glacier frequency that we work with for the power law.

From the distribution of glaciers in 2015, we can determine a power law of the form $N(S) = aS^b$ (with a and b coefficients to be determined) for glacier frequency density by surface area that holds for glaciers not at either extreme end of the area distribution, with the theoretical basis in ref. ¹³, and extending a regional method⁵. In this specific case, we fit the regression line (purple line in Fig. 1a, d) for glaciers between $10^{0.3}$ and $10^{2.6} \text{ km}^2$, with this range selected based on the section of the graph that best fits a straight line while encompassing as many size classes as possible. We note that the regression coefficient is only generated once, globally, rather than for individual regions: this is because the distribution of differently sized glacierized regions is also part of the underlying distribution, and the power-law exponent calculated globally is not necessarily the same as the mean of exponents calculated for each region. Thus, the mass of uncharted glaciers calculated globally is not necessarily the same as the sum of the masses of uncharted glaciers if they were calculated for individual regions. In the context of SLE mass loss contributions, it is important to consider the entirety of the glacierized area rather than focusing on individual regions selected based on geographical convenience.

Estimation of errors. We generate the error values on the regression coefficients by varying the size classes over which the regression is calculated. The upper and lower limits are varied by one size class in each direction, independently, and the resulting distribution of 9 regression coefficients (lower limits of 23rd, 24th and 25th size classes, upper limits of 45th, 46th and 47th size classes in each possible combination) is used as a sample for generation of a 95% confidence interval. The regression coefficient error is generated independently for each of the two regressions performed (one to account for missing glaciers, one to account for disappeared glaciers), which results in a larger proportionate error for disappeared glacier SLE mass loss contribution. The 4.4% proportionate error from the original Marzeion¹¹ model is assumed to hold also for missing and disappeared glaciers and thus added to the error due to the power laws, to determine the total error for these mass loss contributions.

Missing glaciers. In this Letter, 'small glaciers' is taken to mean glaciers with area $10^{-2.0} \text{ km}^2 < S < 10^{0.3} \text{ km}^2$, that is, glaciers below the size of those to which the power law is fitted. We determine a scaling factor for each small glacier size class equal to the ratio of the power-law-predicted frequency density in 2015 to the observed frequency density in 2015 (with a lower cut-off in glacier size for the lower-bound estimate). To obtain an estimate of the missing glacier mass change, the contribution of each RGIv5 small glacier is multiplied by the scaling factor for the 2015 size class it occupies (regardless of what size class the same glacier may have occupied historically). Missing glaciers are defined as those small glaciers that are not included in RGIv5, but are expected to exist in 2015 from this power-law upscaling. The hindcast 1901 glacier distribution with missing glacier scaling applied is shown by the pink line in Fig. 1b, e. The annual GMLSR contribution from missing glaciers is then found by applying the upscaling

factors based on 2015 glacier size class to each glacier's mass-change timeseries for RGIv5 small glaciers.

Disappeared glaciers. We fit the disappeared glacier power law (orange line in Fig. 1c, f) in the same manner as the power law for missing glacier upscaling, but with the pink line (RGIv5 + missing glaciers, hindcast to 1901) as the basis for the calculation of the power-law constant and exponent. However, a correction is needed for the 'bump' in the hindcast RGIv5 + missing glaciers. In large part, we believe this 'bump' to be due to the fact that the glacier model is unable to resolve the merging of two glaciers within a larger valley if they grow (or correspondingly, recombine glaciers, when hindcasting, that were previously the same glacier but split as they shrank); modern separate glaciers in adjoining valleys may have historically been part of a single larger glacier, but the fact that in the 1901 hindcast they are always represented as two separate glaciers artificially inflates some of the smaller size classes in the RGIv5 + missing glacier 1901 distribution, while reducing the glacier count in larger size classes. For this reason, we do not apply any scaling to small glacier size classes for which the power-law-predicted disappeared glacier frequency is lower than the RGIv5 + missing 1901 glacier frequency. The impact of this omission is not expected to be large, and we expect that it results in an overall underestimation of the disappeared glacier SLE mass-loss contribution, because the artificial inflation of size classes below $10^{0.3} \text{ km}^2$ and reduction of larger size classes is expected to result in a smaller power-law exponent. The smaller the power-law exponent (that is, the 'flatter' the distribution for the size classes on which the power law is calculated), the smaller the disappeared glacier mass added through our upscaling.

The time series shown in Fig. 2 (light and dark orange) is not explicitly calculated for disappeared glaciers. Whereas missing glaciers are upscaled on the basis of existing RGIv5 glaciers, disappeared glaciers have no existing analogues for time series upscaling. It is theoretically possible to generate a time series of mass loss by recalculating the power law on a yearly basis and determining how much of the original 1901 disappeared glacier mass is remaining, but the variability in the power-law exponent is almost certain to dominate over actual climate-driven variability. As we know that the contribution is zero in 2015 owing to these glaciers being entirely melted away, we instead show a linear decrease from a maximum in 1901, which is close to what we observe in missing glaciers.

Upper- and lower-bound estimates. We refer to the scaling of all small glacier size classes up to the power laws as the upper-bound contribution estimate. To obtain the lower-bound contribution estimates, we include an additional step: instead of upscaling all small glacier size classes to the calculated power law, we impose a cap on glacier frequency density based on the frequency density at 0.1 km^2 , and upscale only to this cap for size classes between 0.01 km^2 and 0.1 km^2 . This modified upscaling can be seen in Fig. 1d (for missing glaciers) and Fig. 1f (for disappeared glaciers).

In Extended Data Fig. 2, we show the glacier distribution for Switzerland alone, in order to examine the apparent power law for a region where we expect the available glacier inventories to be much more complete. The Swiss Glacier Inventory²² is based on 25-cm-resolution aerial orthophotographs, and as we see in Extended Data Fig. 2, it gives us an apparent power law of exponent 1.16 down to the smallest measured glacier sizes in the RGI. In fact, the RGI itself also exhibits such a power law—with exponent 1.26—across small glaciers, and is apparently no less complete for this region than the SGI. With the limited number of glaciers in the SGI (1,420 in total) and in the RGI restricted to Switzerland, we do not have enough data to determine a power law for larger glaciers, so we are unable to say whether this lower-exponent power law is a characteristic of the region, or a characteristic of the distribution of smaller glaciers with a transition into a steeper power law for larger glaciers. As a compromise, instead of guessing at transitions between power laws for different glacier size scales, we suggest a lower-bound estimate of the uncharted glaciers' contribution by assuming a cut-off at 0.1 km^2 , so the power law observed for larger glaciers continues down to 0.1 km^2 , and below this the distribution is flat. The exponent observed in Switzerland lies somewhere between the lower bound (effectively an exponent of 0) and the upper bound, with the exponent of around 1.8 derived from larger glaciers globally.

For missing glaciers, the upper- and lower-bound estimates are based on the same power law, calculated for RGIv5 glaciers homogenized to 2015, but for disappeared glaciers, because the power law is based on the distribution of RGIv5 + missing glaciers hindcast to 1901 and this differs based on the upscaling used for missing glaciers, the upper- and lower-bound estimates are based on separately calculated power laws, with different exponents. In practice, we note a slightly smaller exponent (but not by much) for the lower-bound estimate, probably owing to the lessened effect of the merging of glaciers when hindcasting caused by the model not accounting for glaciers coming together as their area increases (see 'bump' in Fig. 1c, with a much less noticeable bump in Fig. 1f). We also note that the lower-bound contribution for missing glaciers is by definition smaller than the upper bound, because the upscaling is from the same

base glacier distribution to a strictly smaller-than-pure-power-law distribution. This is not the case for the lower-bound contribution of disappeared glaciers. Owing to the potentially different power-law exponents and the different distributions that are being upscaled from, the lower-bound disappeared-glacier contribution could be larger than the upper-bound contribution, although in practice this is not the case.

Impact of Antarctic peripheral glaciers. Glaciers in the RGIv5 Antarctic and Subantarctic region are unique in this study, as our climate data does not extend to these latitudes. This means that none of the glaciers in this region can be explicitly modelled (Extended Table 1), so global mean mass balance within each size class is assumed for each glacier. This is a strong and not well justified assumption, so it is worthwhile to consider both why it is still valuable to include these glaciers in our analysis, and what the impact on the results is if the region is removed.

Inclusion of Antarctic and Subantarctic peripheral glaciers is desirable, if possible, because the basis for a global upscaling of small glaciers must be a global glacier inventory. Generation of independent power laws on a regional basis and summing the upscaling for uncharted glaciers over these regions does not yield the same results as performing the upscaling based on global glacier distribution, and the distribution of glaciers across regions containing different-sized glacier populations is fundamentally part of the global distribution we are trying to represent. Furthermore, the definition of the RGI regions is largely a matter of convenience, and the ability to artificially partition the world's glaciers into geographically separate boxes does not reflect the fact that the overall distribution of glaciers is the result of the interaction of much less separable factors such as topography, precipitation and surface energy balance, which are variables that are more continuous across glaciated and non-glaciated areas. In the same way that individual glaciers within a region are part of a larger pattern of glaciated area within that region, individual regions are part of a larger pattern of glaciated regions across the globe. Antarctic and Subantarctic peripheral glaciers are part of this global distribution, so we consider their inclusion worthwhile despite the additional modelling assumption, provided they do not have a clearly destabilizing influence on the overall results.

Although the Antarctic and Subantarctic region comprises a large amount of overall glacier mass, it does not represent a large proportion of the total area of small glaciers (4.7%, as compared to 43.1% and 19.3% in the Greenland Periphery and Central Asia regions, respectively). Small glaciers are the ones that contribute to the upscaling for uncharted glaciers, so lack of explicit surface mass balance modelling for Antarctic and Subantarctic peripheral glaciers does not have a larger effect on the upscaled glacier SLE mass-loss contribution. Nevertheless, for completeness we provide data for the global modelling and upscaling with the Antarctic and Subantarctic region removed, corresponding to the same data included in the main text including the region. Only the upper-bound estimate is compared, as the intention is to give an impression of the maximal effect of including or removing Antarctic and Subantarctic glaciers. Removing Antarctic and Subantarctic glaciers, the RGIv5 SLE mass-loss contribution is reduced to 75.8 ± 3.3 mm SLE (from 89.1 ± 3.9 mm including Antarctic and Subantarctic peripheral glaciers), but the missing- and disappeared-glacier contributions actually increase (insignificantly; $P = 0.14$ and 0.50 respectively) to 49.1 ± 5.2 mm and 6.3 ± 2.5 mm SLE respectively (from 42.7 ± 6.5 mm and 5.3 ± 2.4 mm SLE, respectively, including the Antarctic and Subantarctic peripheral glaciers). The reason for the increase in these contributions when Antarctic and Subantarctic glaciers are removed appears to be a change in power-law exponents to -1.83 ± 0.01 for the initial missing-glacier upscaling and -2.01 ± 0.04 for the disappeared-glacier upscaling (from -1.80 ± 0.01 and -1.98 ± 0.04 respectively); these increase the amount of upscaling applied for small glacier size classes by enough to more than account for the reduced overall number of RGIv5 small glaciers used in the dataset, given the aforementioned small proportion of small glaciers that are found in the Antarctic and Subantarctic region.

Exponent constraints. The nature of the power-law explanation of glacier distribution places certain constraints on the power law in order for the outcome to be physically plausible. We concern ourselves with three integrals that relate to the distribution:

$$\int_m^n N(S) dS \quad (1)$$

giving the total number of glaciers with area between m and n

$$\int_m^n SN(S) dS \quad (2)$$

giving the total area of all glaciers with areas between m and n , and

$$\int_m^n jS^k N(S) dS \quad (3)$$

giving the total volume of all glaciers with areas between m and n , with the exponent k and constant j being volume/area scaling factors $k = 1.375$ and $j = (0.0340 \text{ km})^{(3-2k)}$ taken from literature^{23,24}. In our analysis, we find a practical upper bound on the power law of $10^{2.6} \text{ km}^2$, and this is understood to be a physically meaningful upper bound reflecting restrictions on how many large glaciers can exist in glacierized regions of limited size, so we can fix $n = 10^{2.6}$ and do not need to worry about the convergence of these intervals as n increases. At the lower end of glacier areas, we do not have a physically meaningful cutoff for minimum glacier size m . We fix a lower limit of $10^{-2.0} \text{ km}^2$, because this is the smallest glacier size represented in RGIv5, but this is purely a limitation of the dataset, so in order to have physically plausible power laws, we should expect convergence in some of these integrals as m tends to zero. In equations (2) and (3), convergence is necessary as the total area and total volume of glaciers globally must be finite regardless of how small we make our minimum limit on glacier size, but equation (1) should not necessarily converge, because dropping the threshold for what we consider to be glaciers can plausibly add huge numbers of increasingly small ice masses. In essence, we can continue to add large numbers of increasingly small glaciers as long as they do not contribute a substantial overall area or volume. As $N(S)$ is proportional to S^b for the power-law exponent b , in order for equations (2) and (3) to converge as m tends to zero, we require, respectively:

$$b + 1 > -1 \quad (4)$$

$$b + 1.375 > -1 \quad (5)$$

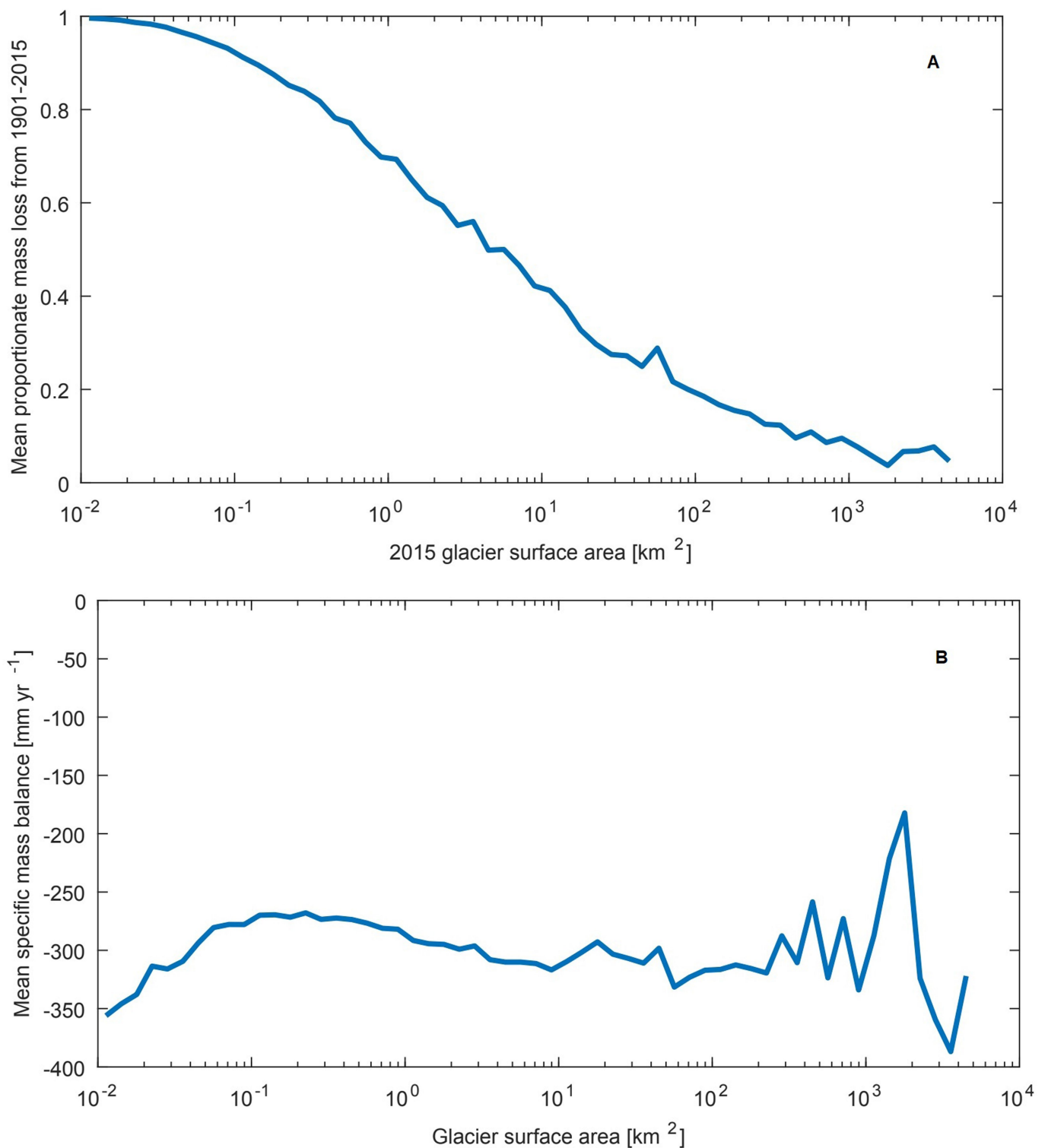
Equation (5) is satisfied comfortably by every value of the exponent for all the power laws we generate, meaning that the total ice volume is at least theoretically consistent in every case. Equation (4) is satisfied comfortably for both the missing glacier power laws with and without Antarctic and Subantarctic glaciers, so the power-law explanation of 2015 glaciers is theoretically consistent. Both with and without Antarctic and Subantarctic glaciers, the upper-bound power laws based on 1901 glacier areas to derive disappeared-glacier contributions have error margins that straddle the $b = -2$ threshold for area convergence, and the corresponding power law for the lower bound is below but extremely close to the $b = -2$ threshold. This means there is uncertainty over whether the power law is too steep to be an accurate description of a possible 1901 glacier distribution if the minimum glacier size approaches zero. However, we do recognize that the inability of the model to account for the fact that separate modern glaciers may actually have been part of the same ice masses in the past when they were larger may 'bunch up' the distribution of smaller glaciers (and we note a slightly smaller exponent for the lower bound, where this effect is lessened). We therefore trust the estimate of the missing-glacier SLE mass-loss contribution more than the disappeared-glacier contribution, but we choose to include the figures as part of a consistent whole given that they originate from the same theoretical basis.

Code availability. The glacier model used¹¹ is available from the corresponding author on reasonable request. The remaining code consists of scripts for data processing that are not provided owing to their relative simplicity.

Data availability

The RGIv5 dataset used for glacier area distribution data are available from GLIMS at <https://www.glims.org/RGI/andolph50.html> with identifier doi:10.7265/N5-RGI-50. The updated glacier model output is available from the corresponding author upon reasonable request. The SGI is described in ref.²² with identifier doi:10.1657/1938-4246-46.4.933, and data was available from the authors on reasonable request. The data generated for this paper is not provided owing to the difficulty of representing a collection of matrices indexed by glacier size class and year in a simple CSV file in a way that is easily readable, but the data is available from the corresponding author on reasonable request.

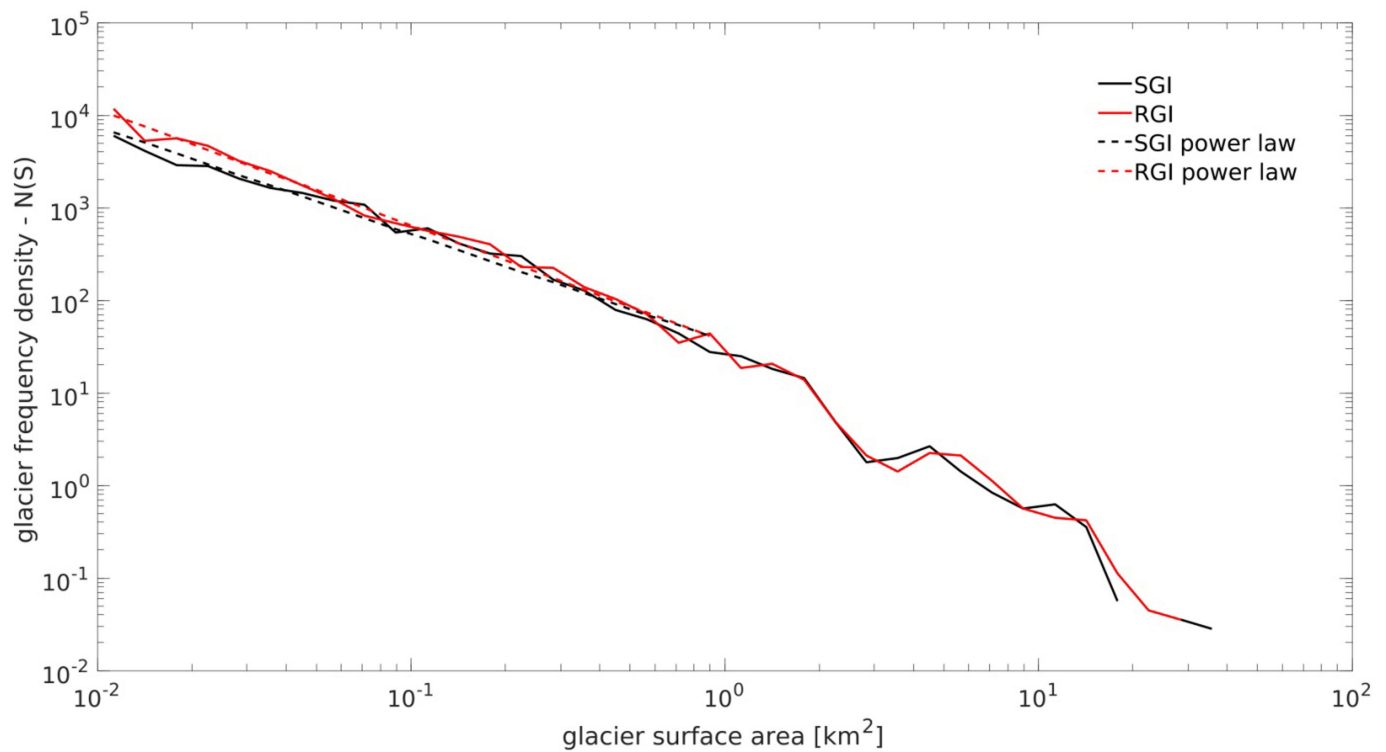
22. Fischer, M., Huss, M., Barboux, C. & Hoelzle, M. The new Swiss Glacier Inventory SGI2010: relevance of using high-resolution source data in areas dominated by very small glaciers. *Arct. Antarct. Alp. Res.* **46**, 933–945 (2014).
23. Bahr, D., Meier, M. & Peckham, S. The physical basis of glacier volume-area scaling. *J. Geophys. Res.* **102**, 20355–20362 (1997).
24. Bahr, D. Global distributions of glacier properties: a stochastic scaling paradigm. *Water Resour. Res.* **33**, 1669–1679 (1997).
25. Fischer, M., Huss, M. & Hoelzle, M. Surface elevation and mass changes of all Swiss glaciers 1980–2010. *Cryosphere* **9**, 525–540 (2015).



Extended Data Fig. 1 | RGIv5 glacier change statistics by size class.

a. Mean specific glacier mass balance by glacier size class. The fact that this graph is relatively flat suggests that the differing mass balance between small glaciers and larger glaciers is not a driver for small glaciers (and by extension missing glaciers) contributing a large amount to SLE mass loss relative to their current ice mass. Glacier size does not strongly affect mean specific mass balance, and this weak dependence is also shown in

observations from the literature²⁵. **b.** Mean proportion of 1901 mass lost between 1901 and 2015 as a function of glacier size class. The smallest glaciers that exist in 2015 typically lost almost all of their 1901 mass, with the proportion dropping consistently as 2015 glacier size increases, up to the largest glaciers in 2015, which have seen an average of less than 10% of their mass disappear since 1901.



Extended Data Fig. 2 | Glacier distribution for Switzerland. We believe that in Switzerland, the RGIv5 (solid red) has a much better representation of small glaciers. The Swiss Glacier Inventory²² (SGI) (solid black), which is based on high-resolution orthophotographs, and which is therefore believed to have better representation of small glaciers

than is available globally, shows good agreement with the RGI. The power laws for the RGI and SGI (dashed red and dashed black respectively) are calculated for the 10^{-2} to 10^0 km^2 range, and show that a credible power law exists in this region down to the smallest glacier sizes, albeit with reduced exponents (1.26 and 1.16 respectively).

Extended Data Table 1 | Distribution of unmodelled glaciers

RGI Region	Percentage of glacierized area that cannot be modeled	Percentage of global small glacier area in this region
Alaska	0.2%	2.2%
Western Canada and US	0.7%	0.1%
Arctic Canada North	3.4%	8.5%
Arctic Canada South	0.7%	3.3%
Greenland Periphery	12.3%	43.1%
Iceland	0.0%	0.0%
Svalbard	6.2%	0.5%
Scandinavia	0.0%	0.0%
Russian Arctic	28.4%	1.4%
North Asia	4.7%	0.5%
Central Europe	2.7%	0.9%
Caucasus and Middle East	14.1%	2.4%
Central Asia	2.3%	19.3%
South Asia West	0.7%	3.6%
South Asia East	0.8%	1.6%
Low Latitudes	15.0%	4.8%
Southern Andes	0.8%	2.9%
New Zealand	0.9%	0.2%
Antarctic and Subantarctic	100.0%	4.7%

Each RGI region may contain glaciers for which the model fails. Antarctic and Subantarctic peripheral glaciers all fail owing to the absence of CRU data for the appropriate latitudes, while in other regions the iteration to find an initial 1901 glacier area fails in a minority of cases. The percentage of glacierized area in each region that cannot be modelled is shown alongside the percentage of global small-glacier area (all glaciers less than $10^{0.3}$ km² in size) in each region (which relates to how much the region affects the upscaling of small glaciers to account for uncharted glaciers). Notably, the Greenland Periphery region contains a disproportionately large amount of the world's total small-glacier area, while the Antarctic and Subantarctic region contains relatively little. At present, we are unable to determine whether this is primarily due to differing regional distributions (for example, Antarctica has more large glaciers but fewer small glaciers than Greenland) or differing data quality.

Dinosaur egg colour had a single evolutionary origin

Jasmina Wiemann^{1*}, Tzu-Ruei Yang² & Mark A. Norell³

Birds are the only living amniotes with coloured eggs^{1–4}, which have long been considered to be an avian innovation^{1,3}. A recent study has demonstrated the presence of both red-brown protoporphyrin IX and blue-green biliverdin⁵—the pigments responsible for all the variation in avian egg colour—in fossilized eggshell of a nonavian dinosaur⁶. This raises the fundamental question of whether modern birds inherited egg colour from their nonavian dinosaur ancestors, or whether egg colour evolved independently multiple times. Here we present a phylogenetic assessment of egg colour in nonavian dinosaurs. We applied high-resolution Raman microspectroscopy to eggshells that represent all of the major clades of dinosaurs, and found that egg colour pigments were preserved in all eumaniraptorans: egg colour had a single evolutionary origin in nonavian theropod dinosaurs. The absence of colour in ornithischian and sauropod eggs represents a true signal rather than a taphonomic artefact. Pigment surface maps revealed that nonavian eumaniraptoran eggs were spotted and speckled, and colour pattern diversity in these eggs approaches that in extant birds, which indicates that reproductive behaviours in nonavian dinosaurs were far more complex than previously known³. Depth profiles demonstrated identical mechanisms of pigment deposition in nonavian and avian dinosaur eggs. Birds were not the first amniotes to produce coloured eggs: as with many other characteristics^{7,8} this is an attribute that evolved deep within the dinosaur tree and long before the spectacular radiation of modern birds.

The huge diversity of avian egg colour⁹ has previously been attributed to the exploration of empty ecological niches after the extinction of nonavian dinosaurs at the terminal Cretaceous event¹. Different nesting environments, as well as nesting behaviours, are thought to influence egg colour^{10–12}. Egg colour may reflect selective pressure as a result of an ecological interaction between the egg producer and an egg predator (camouflage) or parasite (egg recognition). Avian egg colour has previously been shown to react in a plastic fashion to changes in the incubation strategy or climate, or even in mating behaviour^{1,10,12–18}. However, all previously proposed selective factors rely on the fact that the eggs are exposed to the environment^{10,11} and, with scant exception, not buried or covered. More-recent research suggests that egg colour may have co-evolved with (partially) open nesting habits in nonavian dinosaurs⁶ but offers only a single data point of egg colour outside crown birds, in open-nesting oviraptorid dinosaurs. Information on eggshell pigments in a larger sample of nonavian dinosaurs is required to understand the evolution of egg colour.

Both eggshell pigments—biliverdin and protoporphyrin IX—are tetrapyrroles with minor structural differences that affect their chemical properties and their distribution across the eggshell^{19–22}. In contrast to the more hydrophilic biliverdin, which extends deep into the prismatic zone of the eggshell, the more hydrophobic protoporphyrin—which causes spots and speckles—is restricted to the waxy cuticle^{21,23}. The different solubility properties of biliverdin and protoporphyrin appear to be key to their preservation potential^{6,21,22,24}. Protoporphyrin is more resilient to elution than biliverdin but both pigments are preserved in detectable trace amounts^{6,24}. Eggshell pigments appear to be restricted, if not bound, to the proteinaceous scaffold of the eggshell matrix²⁵.

Proteins transform during diagenesis into pyrrole-, pyridine- and imidazole-rich polymers through oxidative crosslinking²⁶; the resulting protein fossilization products (PFPs) appear similar to biliverdin and protoporphyrin IX in their chemical composition. Raman spectroscopy distinguishes between true egg-colour pigments and pigment-like PFPs²⁶ (Extended Data Fig. 1), and identifies and maps out pigments over eggshell surfaces and across vertical egg sections to characterize colour patterns and pigment deposition in fossil eggs. Placing this information in a phylogenetic context offers insights into whether egg colour evolved once within nonavian dinosaurs or multiple times independently, and might help to identify selective factors.

In our sample of nineteen archosaur eggshells, egg colour pigments are absent in eggshells of *Alligator mississippiensis*, the North American hadrosaurid *Maiaasaura peeblesorum*, the South American saltasaurid, the French titanosaurid and the North American troodontid (Fig. 1, Extended Data Figs. 2, 3). Egg colour pigments are preserved in eggshells from the oviraptorid *Heyuannia huangi*, Mongolian microtroodontids, the Chinese and Mongolian troodontids, the dromaeosaurid *Deinonychus antirrhopus*, the Mongolian enantiornithine, *Psammornis rothschildi*, *Rhea americana*, the North American ratite, *Dromaius novaehollandiae* and *Gallus domesticus* (Fig. 1, Extended Data Figs. 2, 3).

Only biliverdin was detected in *D. novaehollandiae*, whereas only protoporphyrin IX was present in the eggshells of the Mongolian microtroodontid (MAE 14-40 (specimen codes in parentheses)), the Chinese and Mongolian troodontids, the Mongolian enantiornithine, *P. rothschildi* and *G. domesticus*. Both egg colour pigments were detected in eggshells from *H. huangi*, the Mongolian microtroodontid (IGM 100/1323) and macrotroodontid (AMNH FARB 6631), *D. antirrhopus*, *R. americana* and the North American ratite. The presence of eggshell pigments corresponds to (partially) open nesting habits (Fig. 1).

All eggshell and associated sediment samples were plotted on a whole spectra-based principal component analysis (PCA) (Extended Data Fig. 4 and its Source Data). Principal component 1 (PC1, 57.118%) represents variability in pigment type, concentration and mode of eggshell alteration, whereas principal component 2 (PC2, 23.841%) separates samples into unpigmented and pigmented eggshells (Extended Data Fig. 5). The PCA (Extended Data Fig. 4a) revealed that eggshell biomolecules are distinct from organic material in the sediment, with both clusters separating across PC1 (73.116%). Within the eggshell cluster, extant and fossil materials are separated across PC2 (10.977%) (Extended Data Fig. 4a). A separate PCA (Extended Data Fig. 4b) based on the spectral fingerprint region of biliverdin and protoporphyrin IX (1,500 cm^{−1}–1,650 cm^{−1} ± 2 cm^{−1}) included all fossil eggshell samples: the resulting chemo-space identified a characteristic cluster of pigmented eggshells, distinct from a separate cluster of unpigmented eggshells. Mapping protoporphyrin IX on the eggshell surface (Fig. 2a) demonstrated that the eggs of *H. huangi* were spotted, as were those of the Mongolian microtroodontids and troodontids, *D. antirrhopus*, and the Mongolian enantiornithine. Reconstructions of the egg colours are shown in Fig. 2a.

Depth profiles (Fig. 2b) across vertical eggshell sections show that pigments are absent in all layers of the *A. mississippiensis* eggshell as

¹Department of Geology & Geophysics, Yale University, New Haven, CT, USA. ²Steinmann Institute for Geology, Mineralogy, and Paleontology, University of Bonn, Bonn, Germany. ³Division of Vertebrate Paleontology, American Museum of Natural History, New York, NY, USA. *e-mail: jasmina.wiemann@yale.edu

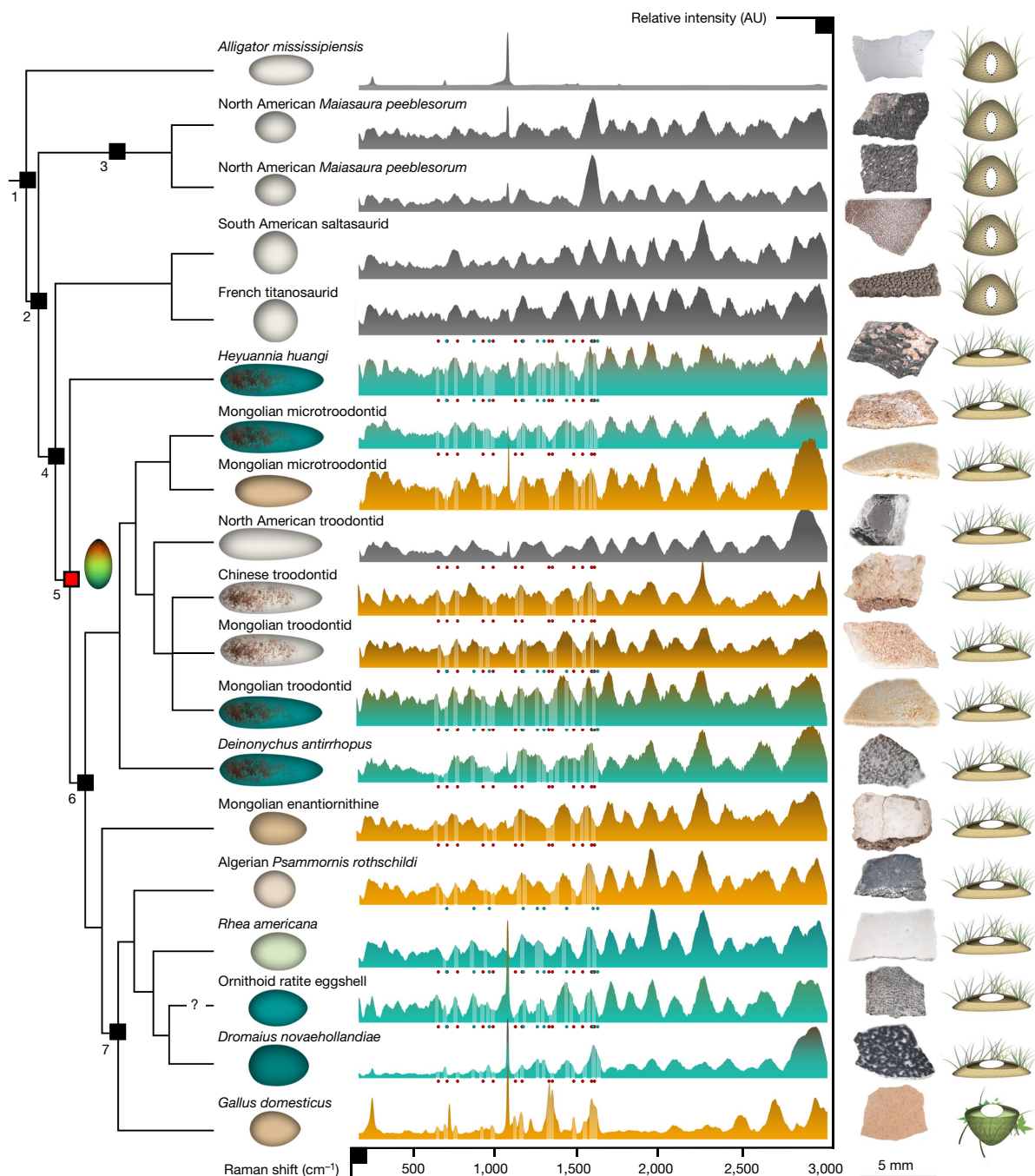


Fig. 1 | Stacked Raman spectra and ancestral state reconstruction on a pruned supertree. The Raman spectra represent bands at $200\text{--}3,000\text{ cm}^{-1} \pm 2\text{ cm}^{-1}$, from 6 accumulations and 20 s of exposure; spectra are baselined and normalized. Ancestral state reconstruction is based on published phylogenies^{8,28,29} ($n = 19$ sampled taxa), and maximum parsimony. The internal nodes are (1) Archosauria, (2) Dinosauria, (3) Ornithischia, (4) Saurischia, (5) Eumaniraptora, (6) Paraves and (7) Aves. The egg icon in the phylogeny labels Eumaniraptora. All terminal

taxa are represented by an icon indicating egg shape, and an example of reconstructed colour. If pigments are present, the area below the spectral function is coloured in blue (biliverdin) or orange (protoporphyrin IX), and all pigment bands are labelled with either blue (biliverdin) or red (protoporphyrin IX) dots. Relative Raman band intensities may vary owing to differential preservation. Photographs show the samples and nest icons encode three nesting strategies: buried, (partially) open ground and open tree nesting. AU, arbitrary units.

well as on the surface. The pigment stratification in *D. novaehollandiae*, *D. antirrhopus*, the Mongolian troodontid (AMNH FARB 6631) and *H. huangi* appears almost identical: in these samples, both pigments show a peak in concentration in the eggshell cuticle and increased concentration values through the entire prismatic zone. In the eggshells of *G. domesticus* and the Mongolian enantiornithine, the protoporphyrin IX signal reaches a peak in the organic and mineralized cuticle and extends into the uppermost prismatic zone. This is consistent with the absence of biliverdin in

the high-resolution point measurements of these samples. Only the Mongolian troodontid (AMNH FARB 6631) produced a different signal in the depth profiles to those obtained in the point measurements: in addition to very weak bands assigned to protoporphyrin IX in the point measurements for this eggshell, the depth profile revealed a weak biliverdin signal that is apparently restricted to the deeper eggshell layers. In contrast to the depth profiles in fresh *D. novaehollandiae* and *G. domesticus* eggshells, all fossil samples showed minor evidence of pigment elution (Fig. 2b). Only the Mongolian

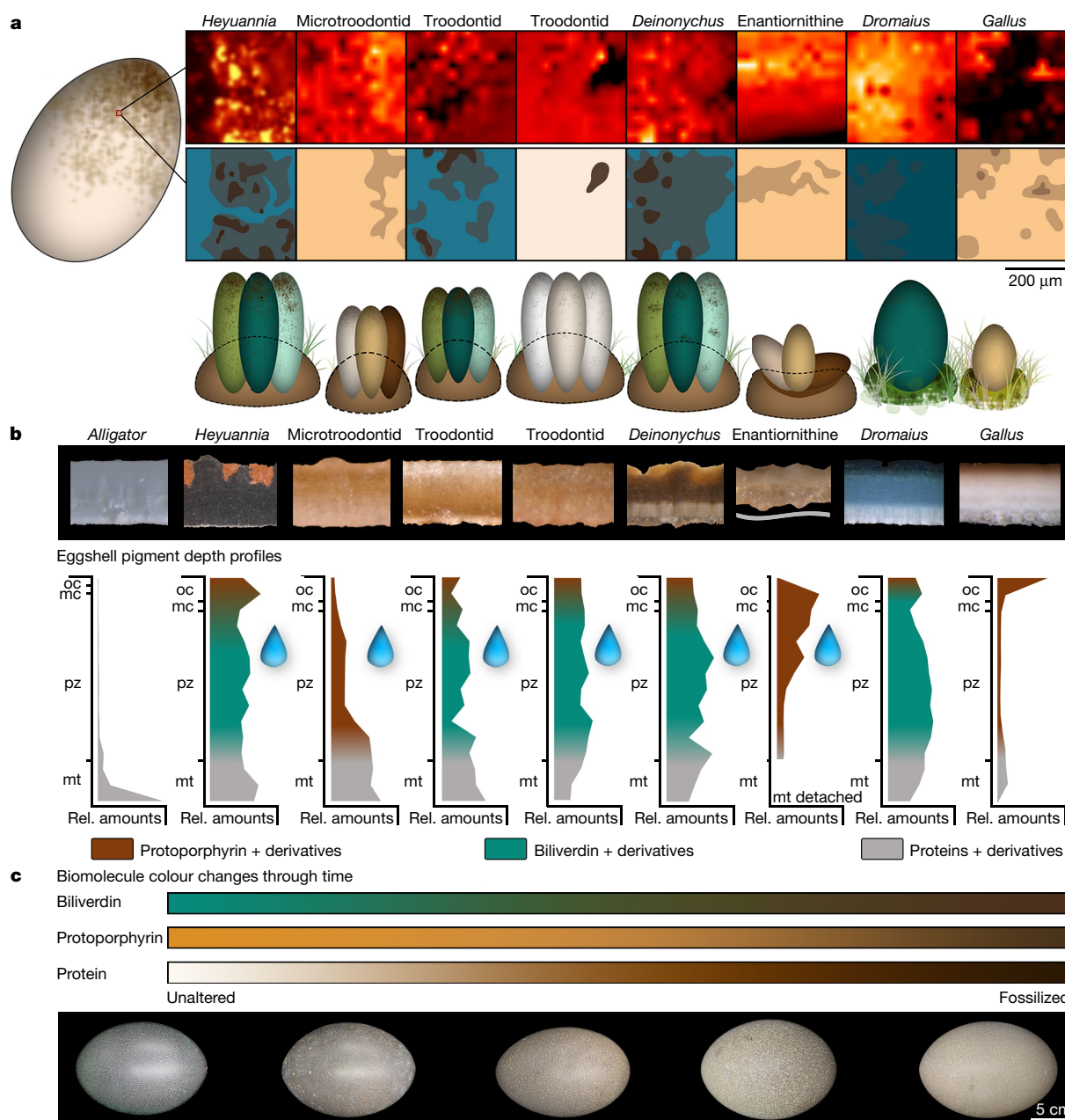


Fig. 2 | Egg colour reconstruction. **a**, Top, eggshell-pigment surface maps. $n = 8$; selection criteria were pigment presence (Fig. 1) and sufficient surface exposure. Protoporphyrin was mapped ($1,350 \text{ cm}^{-1} \pm 2 \text{ cm}^{-1}$, 2 accumulations, 5 s of exposure) with three independent repetitions, which yielded similar results. Increased signal intensity (yellow) is relative to the lowest signal intensity (black, which equals background noise in the absence of protoporphyrin IX). Bottom, egg reconstructions that combine information from panels above, **b** and Fig. 1 into a range of potential colours for the fossil eggshells. From left to right: *H. huangi*, Mongolian microtroodontid (MAE 14-40), Mongolian troodontid (AMNH FARB 6631), Mongolian troodontid (IGM 100/1003), *D. antirrhopus*, Mongolian enantiornithine, *D. novaehollandiae* and *G. domesticus*. **b**, Pigment depth

profiles across vertical sections of eggs from *A. mississippiensis*, plus the taxa shown in **a** ($n = 9$ specimens). Depth profiles were repeated three times independently, which yielded similar results. Photographs and depth profiles are not at the same scale. The distribution of protoporphyrin IX (red), biliverdin (blue) and proteinaceous matter (grey) is based on Raman point measurements and line maps ($1,166 \text{ cm}^{-1} \pm 2 \text{ cm}^{-1}$). Droplet icons indicate pigment elution. oc, organic cuticle; mc, mineralized cuticle; pz, prismatic zone; mt, membrana testacea. **c**, Visualization of gradual colour change of eggshell pigments and proteinaceous matter through time, based on observations of eggshells (*D. novaehollandiae* and *Casuarus casuaris*) and on a previous study²⁶.

microtroodontid appeared to have experienced very intense pigment elution, as can be seen by comparing its depth profile to that of *G. domesticus* (protoporphyrin IX only, Fig. 2b). No trace of eggshell pigments was found in any of the sediment samples that surrounded the eggshells (Extended Data Fig. 6).

We infer endogeneity of the eggshell pigments detected on the basis of a number of criteria: the required 22 Raman band matches for biliverdin and protoporphyrin IX (Extended Data Fig. 1), the statistically supported separation of unpigmented and pigmented fossil eggshells

(Extended Data Fig. 4b), the exclusion of pigment-like PFPs in the Raman-band identification (Extended Data Fig. 1), the similarity between pigment depth profiles in fresh and fossil eggshells (Fig. 2b) and the absence of pigments in the host sediments (Extended Data Figs. 4a, 6).

The phylogenetic distribution of egg colour within this archosaur sample supports a single evolutionary origin within nonavian theropod dinosaurs (Fig. 1). The homology between nonavian dinosaur and bird eggshell colour is further supported by an identical mode of eggshell

pigment deposition. This is based on a comparison of the depth profiles of eggshell pigments in nonavian eumaniraptorans with those of *D. novaehollandiae* and *G. domesticus*, which reveals that the stratification of pigments appears almost identical (Fig. 2) both among taxa that incorporate biliverdin and protoporphyrin IX and among taxa that incorporate only protoporphyrin IX. Only a larger sample of eggshell from taxa deeper in the tree will answer the question of whether egg colour is a true eumaniraptoran synapomorphy, or whether it evolved in more-basal theropods.

The North American troodontid is the only sample within eumaniraptorans⁸ that did not yield any colour signal. Although ornithischian *M. peeblesorum* eggshells from the same locality in the Two Medicine Formation also lacked colour, ornithoid ratite eggshells from comparable beds preserve traces of both biliverdin and protoporphyrin. We therefore interpret the absence of egg colour pigments in these troodontid eggshells (YPM VP PU 023259) as real, rather than taphonomic. Secondary reduction of egg colour is well-documented in modern birds and appears to be associated with increased sunlight exposure—an adaptation against overheating as seen, for example, in ostriches—nocturnality or cave breeding¹².

Our reconstructed egg colours (Fig. 2a) reflect the different preservation potential of biliverdin and protoporphyrin IX. We combined data from the high-resolution Raman point measurements, pigment surface maps and depth profiles of eggshell pigments (Fig. 2b). Because biliverdin is more likely to be washed out than the less-hydrophilic protoporphyrin IX, preserved traces of biliverdin indicate higher original concentrations⁶. Signs of elution are evident in all of our fossil eggshell samples, even though associated sediments lack detectable amounts of washed-out pigments (Extended Data Figs. 4a, 6). PFPs of the organic scaffold inside the eggshell produce a brownish-black colour²⁶ and original, unaltered eggshell pigments survive as traces through deep time⁶. It appears that part of the pigment fraction undergoes chemical alteration that contributes to a brownish discolouration (Fig. 2c).

The oviraptorid *H. huangi* belongs to the basalmost clade with coloured eggs (Fig. 1) in this study, and incorporates both biliverdin and protoporphyrin IX. It also represents the basalmost nonavian theropod group that built open nests—a transition that occurred at the base of Eumaniraptora and is found in nearly all its descendants^{6,27}. Absence of egg colour pigments in all taxa burying their eggs corroborates the hypothesis⁶ that egg colour co-evolved with open nesting habits (Fig. 1).

Egg colour is homologous in nonavian and avian dinosaur, and can be traced back to a single evolutionary origin in eumaniraptorans. The reconstructed diversity in egg colour and pattern (Fig. 2a) among nonavian dinosaurs (Fig. 1) mirrors that in extant birds^{1,3,5,18}. The evolution of spots and speckles in the eggshells of modern birds is often associated with individual recognition strategies in communally nesting groups, and intensively nest-parasitized host species¹⁸. Individual patterning of nonavian theropod eggs suggests the presence of much more complex nesting niches than previously recognized. Birds inherited a powerful molecular toolkit that enables them to colour their eggs; this discovery demands re-evaluation of evolutionary trends in egg colour at the base of modern birds.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0646-5>.

Received: 6 March 2018; Accepted: 24 August 2018;

Published online 31 October 2018.

1. Kilner, R. M. The evolution of egg colour and patterning in birds. *Biol. Rev. Camb. Philos. Soc.* **81**, 383–406 (2006).
2. Cherry, M. I. & Gosler, A. G. Avian eggshell coloration: new perspectives on adaptive explanations. *Biol. J. Linn. Soc.* **100**, 753–762 (2010).
3. Cassey, P. et al. Why are birds' eggs colourful? Eggshell pigments co-vary with life-history and nesting ecology among British breeding non-passerine birds. *Biol. J. Linn. Soc.* **106**, 657–672 (2012).

4. Packard, M. J. & Seymour, R. S. in *Amniote Origins: Completing the Transition to Land* (eds Sumida, S. S. & Martin, K. L. M.) Ch. 8, 265–290 (Academic, San Diego, 1997).
5. Kennedy, G. Y. & Vevers, H. G. A survey of avian eggshell pigments. *Comp. Biochem. Physiol. B* **55**, 117–123 (1976).
6. Wiemann, J. et al. Dinosaur origin of egg color: oviraptors laid blue-green eggs. *PeerJ* **5**, e3706 (2017).
7. Norell, M. A., Clark, J. M., Chiappe, L. M. & Dashzeveg, D. A nesting dinosaur. *Nature* **378**, 774–776 (1995).
8. Turner, A. H., Makovicky, P. J., & Norell, M. A. A Review of dromaeosaurid systematics and paravian phylogeny. *Bull. Am. Mus. Nat. Hist.* **371**, 1–206 (2012).
9. Stoddard, M. C. & Prum, R. O. How colorful are birds? Evolution of the avian plumage color gamut. *Behav. Ecol.* **22**, 1042–1052 (2011).
10. Komdeur, J. & Kats, R. K. Predation risk affects trade-off between nest guarding and foraging in Seychelles warblers. *Behav. Ecol.* **10**, 648–658 (1999).
11. Gillis, H., Gauffre, B., Huot, R. & Bretagnolle, V. Vegetation height and egg coloration differentially affect predation rate and overheating risk: an experimental test mimicking a ground-nesting bird. *Can. J. Zool.* **90**, 694–703 (2012).
12. Hewitson, W. C. *Eggs of British Birds* (John Van Voorst, London, 1846).
13. Wallace, A. R. *Darwinism: An Exposition of the Theory of Natural Selection with Some of its Applications* (Macmillan, London, 1890).
14. Stoddard, M. C. et al. Camouflage and clutch survival in plovers and terns. *Sci. Rep.* **6**, 32059 (2016).
15. Newton, A. V. *A Dictionary of Birds* (A&C Black, London, 1896).
16. Ishikawa, S. et al. Photodynamic antimicrobial activity of avian eggshell pigments. *FEBS Lett.* **584**, 770–774 (2010).
17. Lahti, D. Population differentiation and rapid evolution of egg colour in accordance with solar radiation. *Auk* **125**, 796–802 (2008).
18. Gosler, A. G., Higham, J. P. & Reynolds, S. J. Why are birds' eggs speckled? *Ecol. Lett.* **8**, 1105–1113 (2005).
19. Ryter, S. W. & Tyrrell, R. M. The heme synthesis and degradation pathways: role in oxidant sensitivity. Heme oxygenase has both pro- and antioxidant properties. *Free Radic. Biol. Med.* **28**, 289–309 (2000).
20. Falk, J. E. *Porphyrins and Metalloporphyrins* (Elsevier, Amsterdam, 1964).
21. Gorchein, A., Lim, C. K. & Cassey, P. Extraction and analysis of colourful eggshell pigments using HPLC and HPLC/electrospray ionization tandem mass spectrometry. *Biomed. Chromatogr.* **23**, 602–606 (2009).
22. Milgrom, L. R. & Warren, M. J. in *The Colours of Life: An Introduction to the Chemistry of Porphyrins and Related Compounds* (ed. Milgrom, L. R.) Ch. 1–5, 1–175 (Oxford Univ. Press, Oxford, 1997).
23. Wang, X. T. et al. Comparison of the total amount of eggshell pigments in Dongxiang brown-shelled eggs and Dongxiang blue-shelled eggs. *Poult. Sci.* **88**, 1735–1739 (2009).
24. Igic, B. et al. Detecting pigments from colourful eggshells of extinct birds. *Chemoecology* **20**, 43–48 (2010).
25. Nys, Y., Gautron, J., Garcia-Ruiz, J. M. & Hincke, M. T. Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. *C. R. Palevol* **3**, 549–562 (2004).
26. Wiemann, J. et al. Fossilization transforms proteins into N-heterocyclic polymers. *Nat. Commun.* <https://doi.org/10.1038/s41467-018-07013-3> (2018).
27. Varricchio, D. J. & Jackson, F. D. Reproduction in Mesozoic birds and evolution of the modern avian reproductive mode. *Auk* **133**, 654–684 (2016).
28. Sereno, P. C. The evolution of dinosaurs. *Science* **284**, 2137–2147 (1999).
29. Pisani, D., Yates, A. M., Langer, M. C. & Benton, M. J. A genus-level supertree of the Dinosauria. *Proc. R. Soc. Lond. B* **269**, 915–921 (2002).

Acknowledgements We thank D. E. G. Briggs for advice and assistance with the manuscript. M. Fabbri, N. Mongiardino Koch, J. Gauthier, K. Zykowski and R. Prum made helpful suggestions. Y.-N. Cheng, Y.-F. Shiao, X. Wu, T. Töpfer, P. M. Sander and K. Zykowski provided eggshell specimens. This research was supported by the Steven Cohen Award of the Society of Vertebrate Paleontology (J.W.), the Macaulay Family Endowment (M.A.N.) and the Division of Paleontology, American Museum of Natural History.

Reviewer information Nature thanks D. Zelenitsky and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.W., T.-R.Y. and M.A.N. designed the research. J.W. designed and performed the experiments. M.A.N. and T.-R.Y. contributed materials. J.W. created the figures. J.W. and M.A.N. wrote the manuscript, which was reviewed by all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0646-5>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0646-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.W.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

We were able to rule out false-negative eggshell pigment assignments owing to taphonomic loss by including different available eggshell ($n = 19$) and sediment ($n = 12$) specimens from similar or identical host rocks. Every eggshell sample was blanked against its host-rock sediment to exclude false-positive results due to potential contamination (see Supplementary Information, chapters 1b–d, 3). The eggshell specimens cover extant *A. mississippiensis*, all major clades of extinct nonavian dinosaurs with multiple representatives, as well as extinct and extant palaeognath and neognath birds (Supplementary Table 2), and were taxonomically identified on the basis of their association with diagnostic adult, juvenile or embryonic skeletal remains or based on a combination of locality, diagnostic eggshell surface ornamentation and microstructure^{30,31} (Supplementary Information, chapter 2).

Eggshell and sediment samples (Supplementary Information, chapter 3) were imaged with a Leica MZ16 dissecting microscope (Optronics camera, LAS Core Software), surface-cleaned with ethanol and subjected to in situ Raman microspectroscopy. Raman spectroscopy was performed using a Horiba800 LabRam with 532-nm excitation (20 mW at the sample surface). The scattered Raman light was detected by an electron multiplier charged-coupled device after being dispersed with an 1,800-grooves-per-mm grating, and passed through a 100- μm entrance slit (hole size, 300 μm). The spectrometer was calibrated using the first-order Si band at 520.7 cm^{-1} . For point measurements (Fig. 1, Extended Data Fig. 1), 6 spectra were accumulated in the 300–2,000 cm^{-1} region for 20-s exposure time each. Fossil soft tissues containing PFPs were obtained from *Allosaurus fragilis* bone (YPM 48) and subjected to the same Raman point-measurement routine, to distinguish tetrapyrrole pigments from similar pigment-free PFPs (Extended Data Fig. 1). This process yielded 22 Raman bands (13 for protoporphyrin IX and 9 for biliverdin) that are diagnostic for egg colour pigments—including previously described pigment bands³²—in a potential PFP environment (Extended Data Fig. 1, Supplementary Information, chapter 1c, d). Further distinction of pigments from PFPs was aided by eggshell net enrichment plots (Extended Data Fig. 3). The eggshell pigments were mapped across eggshell surfaces and through vertical eggshell sections to image potential colour patterns, and to compare modes

of pigment deposition among nonavian and avian dinosaurs. For pigment surface maps (Fig. 2) (protoporphyrin IX only, 1,350 cm^{-1}) and eggshell-pigment depth profiles (1,166 cm^{-1}), 2 spectra were accumulated over 5-s exposure time. Bands that represent protoporphyrin IX (at 1,350 cm^{-1} ; for surface mapping) and both pigments (at 1,166 cm^{-1} ; for depth profiling) were selected for their diagnostic molecular contents and signal intensity under the selected Raman parameters. The spectra were processed in LabSpec 5 (spectral acquisition, mapping acquisition and standard spike removal) and subjected to a standard base-line correction (adaptive baseline, 50%, no offset and no smoothing) and standard normalization (based on the highest band in each spectrum) in SpectraGryph 1.2 spectroscopic software. Reconstructions of egg colours and patterns (Fig. 2a) represent combined data from point measurements, maps and depth profiles.

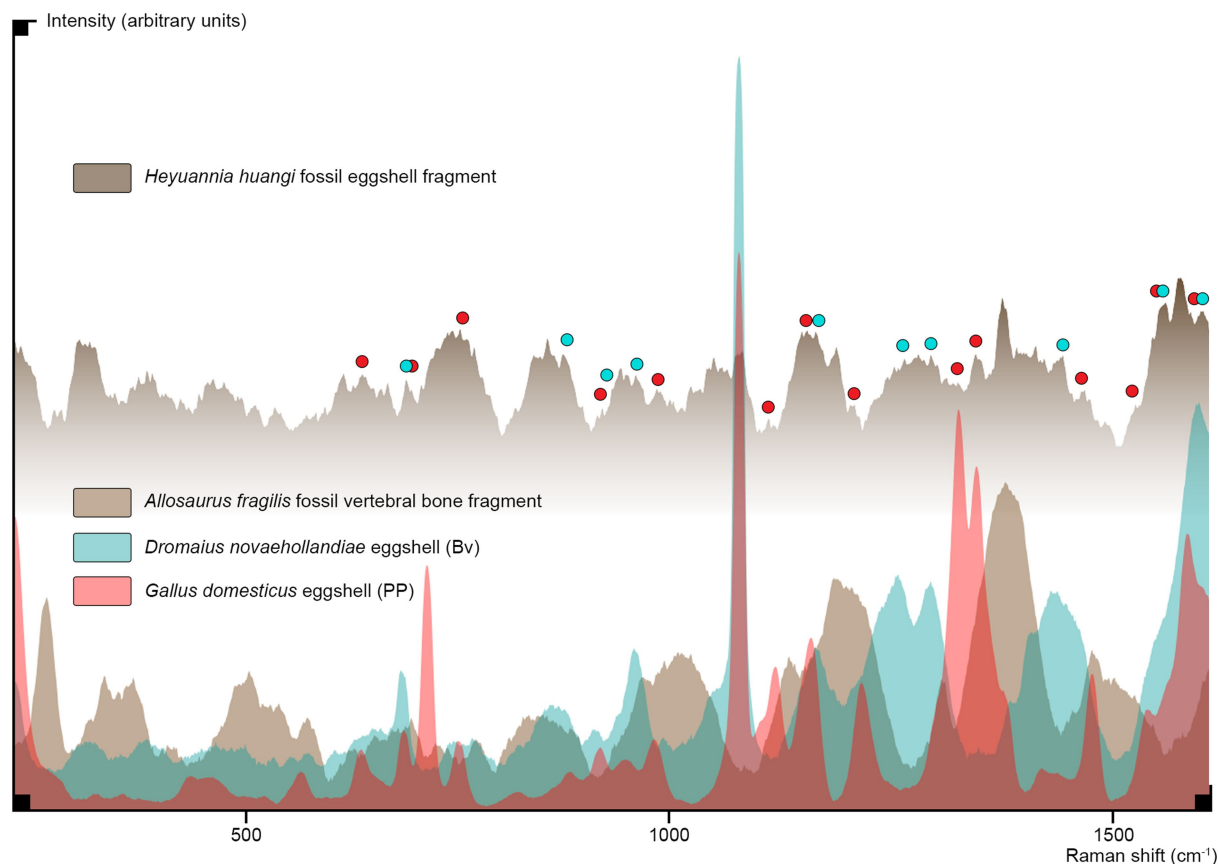
On the basis of the pigment data obtained for nonavian and avian dinosaurs, we performed a parsimony-based character tracing across our manually re-drawn composite phylogeny^{8,28,29} (pruned tree) in Mesquite 3.40 (coding: pigments absent = 0, pigment present = 1) (Extended Data Figs. 4, 6, Supplementary Information, chapter 4). Whole-spectra data of all eggshell and sediment samples (Extended Data Fig. 1), as well as extracted spectral data covering the pigment fingerprint region for fossil eggshells (Extended Data Fig. 1), were transformed into variance–covariance matrices (Extended Data Fig. 4 and its Source Data) and subjected to a PCA, which was performed in Past 3. Our study did not involve experiments that would require sample randomization or blinding of the investigators.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

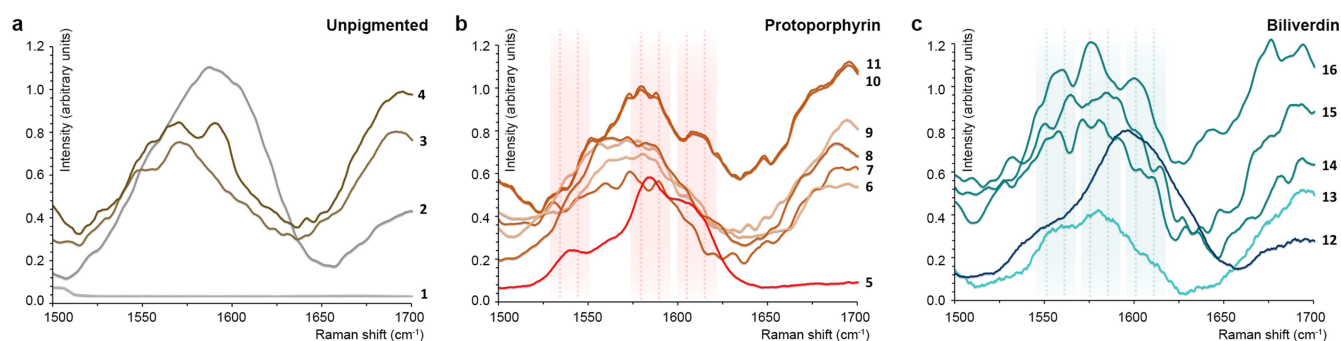
The authors declare that all Raman data supporting the findings of this study are available within the paper (pigment maps and depth profiles), and its Supplementary Information and Source Data.

30. Carpenter, K. et al. (eds) *Dinosaur Eggs and Babies* (Cambridge Univ. Press, Cambridge, 1996).
31. Mikhailov, K. E., Bray, E. S. & Hirsch, K. F. Parataxonomy of fossil egg remains (Vetervata): principles and applications. *J. Vertebr. Paleontol.* **16**, 763–769 (1996).
32. Thomas, D. B. et al. Analysing avian eggshell pigments with Raman spectroscopy. *J. Exp. Biol.* **218**, 2670–2674 (2015).



Extended Data Fig. 1 | Spectral plots that validate the Raman bands indicative of biliverdin and protoporphyrin IX relative to PFPs, based on three representative samples. Spectral plots are $200\text{--}1,600\text{ cm}^{-1} \pm 2\text{ cm}^{-1}$, from 6 accumulations; spectra are baselined and normalized. Each spectrum was repeated three times independently, and yielded similar results. The lowermost spectral plot depicts blue spectra for biliverdin (from *D. novaehollandiae* eggshell), red spectra for protoporphyrin IX (from *G. domesticus*) and brown spectra for fossil

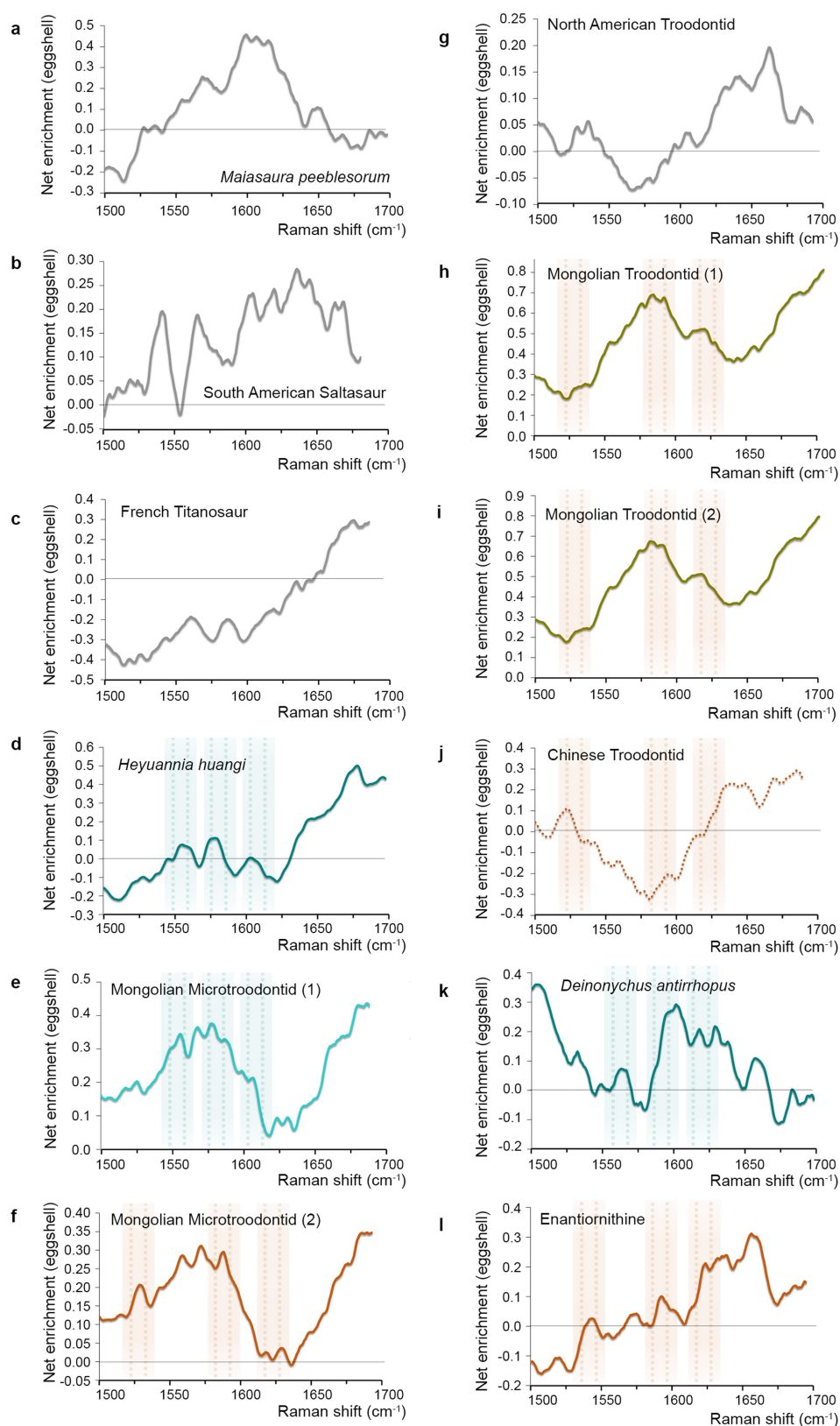
proteinaceous soft tissue (from *A. fragilis* bone; see previous study²⁶). The uppermost dark-brown spectra represent an in situ measurement for the *H. huangi* (NMNS CYN-2004-DINO-05) eggshell that is known to preserve both pigments (biliverdin and protoporphyrin IX). Blue asterisks label bands of biliverdin that differ from PFP; red asterisks label bands of protoporphyrin IX that are absent in fossil soft-tissue remains. Overall, 22 Raman bands are identified that can be generated only by original unaltered pigments, and not by PFPs that are simultaneously present.



Extended Data Fig. 2 | Spectral close-ups of the pigment fingerprint region for nineteen archosaur eggshells. Spectra are at $1,500\text{--}1,700\text{ cm}^{-1} \pm 2\text{ cm}^{-1}$ from 6 accumulations, and are baselined and normalized. **a**, Unpigmented eggshell samples (Fig. 1). **b**, Eggshell samples containing only protoporphyrin IX. Red bands label the set of bands that is diagnostic for protoporphyrin IX eggshell pigment. **c**, Biliverdin-rich eggshell samples. Blue bands label the set of bands that is diagnostic for biliverdin eggshell pigment. Note the characteristic spectral

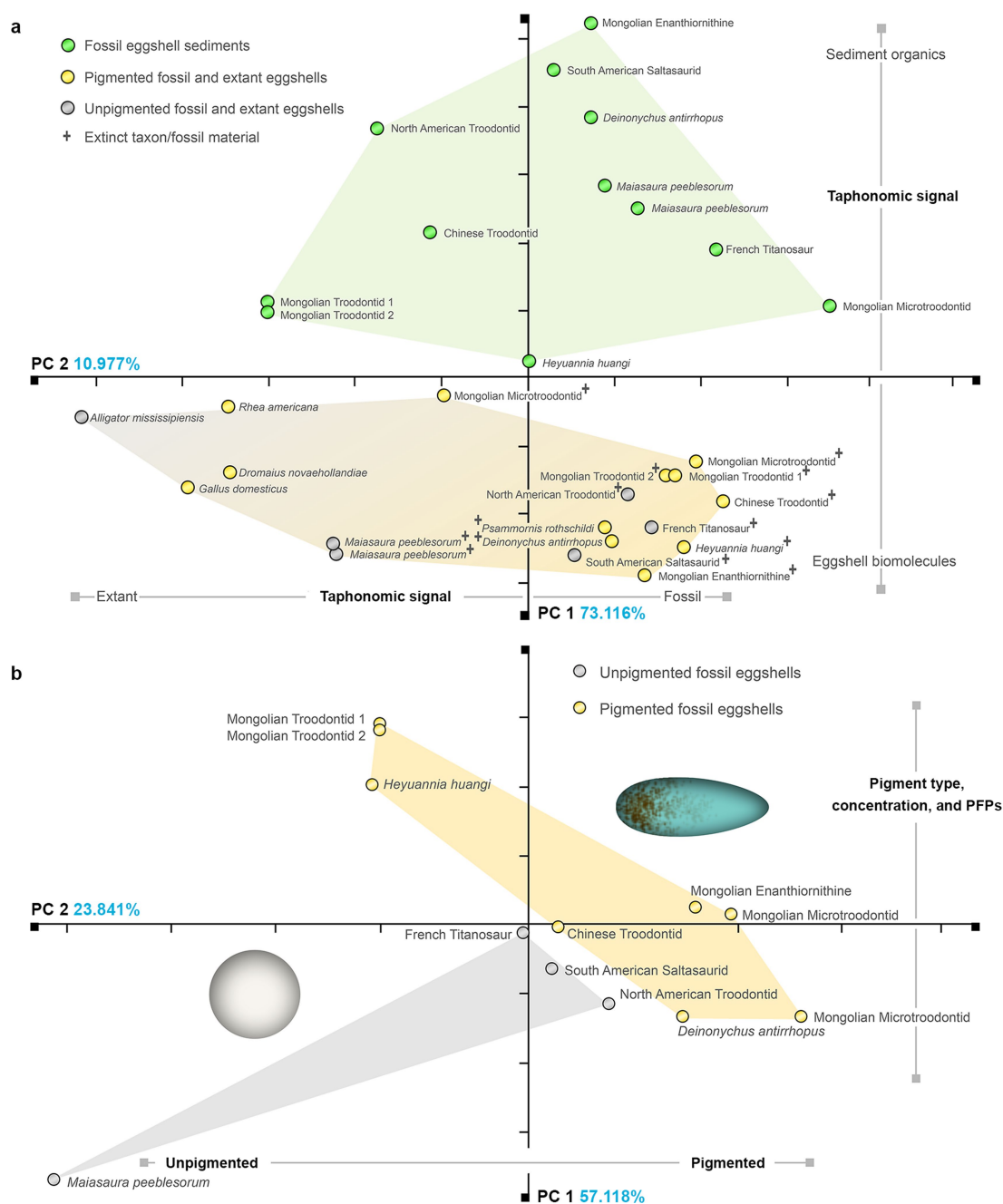
differences between unpigmented and pigmented eggshell samples.

1, *A. mississippiensis*; 2, *M. peeblesorum* (YPM VP PU 023464); 3, South American saltasaurid; 4, French titanosaurid; 5, *G. domesticus*; 6, Chinese troodontid; 7, Mongolian microtroodontid (MAE 14-40); 8, *P. rothschildi*; 9, Mongolian enantiornithine; 10, Mongolian troodontid (AMNH FARB 6631); 11, Mongolian troodontid (IGM 100/1003); 12, *D. noveahollandiae*; 13, *R. americana*; 14, *D. antirrhopus*; 15, Mongolian microtroodontid (IGM 100/1323); 16, *H. huangi*.



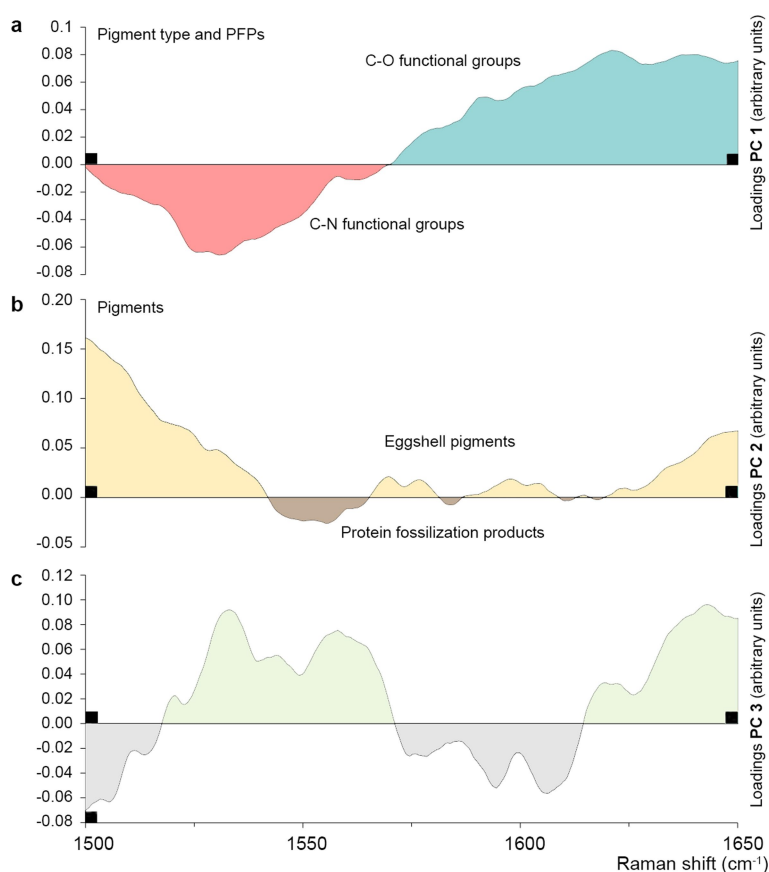
Extended Data Fig. 3 | Eggshell net enrichment plots resulting from subtraction of the sediment spectral functions from their corresponding eggshell spectral functions. Spectra are at $1,500\text{--}1,700\text{ cm}^{-1} \pm 2\text{ cm}^{-1}$ from 6 accumulations, and are base-lined and normalized. Positive values indicate a net enrichment of fossil eggshells in organic compounds relative to their sediment surroundings, whereas negative values indicate a net depletion. Red bands are diagnostic for

protoporphyrin IX, and blue bands are diagnostic for biliverdin. **a**, *M. peeblesorum* (YPM VP PU 023464). **b**, South American saltasaurid. **c**, French titanosaurid. **d**, *H. huangi*. **e**, Mongolian microtroodontid (IGM 100/1323). **f**, Mongolian microtroodontid (MAE 14-40). **g**, North American troodontid. **h**, Mongolian troodontid (AMNH FARB 6631). **i**, Mongolian troodontid (IGM 100/1003). **j**, Chinese troodontid. **k**, *D. antirrhopus*. **l**, Mongolian enantiornithine.



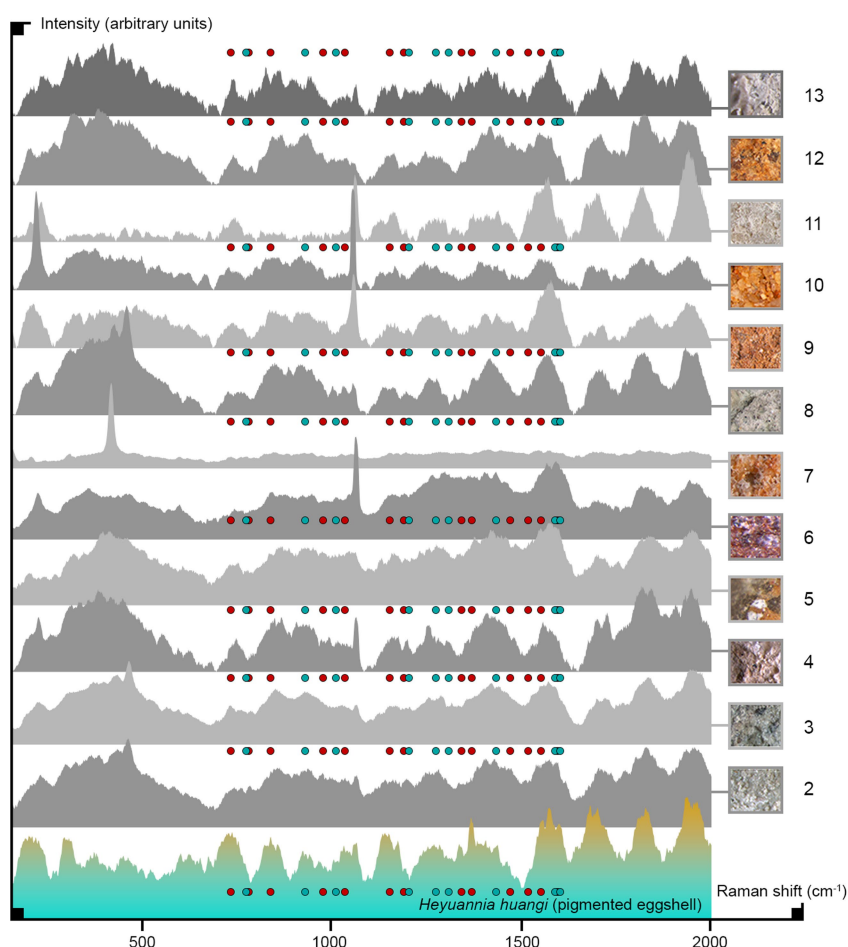
Extended Data Fig. 4 | PCA of spectral data. a, Whole spectra-based PCA of all modern ($n = 4$ biologically independent samples) and fossil ($n = 14$ biologically independent samples) eggshell and sediment samples ($n = 12$ independent samples). Sediment samples cluster together (green), and do not overlap the cluster of eggshell samples (grey-to-yellow). Eggshell biomolecules and sediment organics are distinct. Within the eggshell cluster, extant material is separated from fossil material. The proximity of unpigmented, protoporphyrin IX-bearing and biliverdin-rich extant eggshell material emphasizes that the taphonomic signal overprints the pigment signal at the level of a whole spectrum-based PCA. PC1 (73.116%) and PC2 (10.977%) are characterized by high support values and explain most of the variation in the eggshell and sediment sample set. **b**, PCA based on the pigment fingerprint region (1,500–1,650 $\text{cm}^{-1} \pm 2 \text{ cm}^{-1}$) of protoporphyrin IX and biliverdin in

fossil eggshell material ($n = 12$ biologically independent samples). The exclusion of fresh and 'subfossil' (*P. rothschildi* and ornithoid ratite) eggshell material and the extraction of the pigment fingerprint region enable a chemo-space separation based on the eggshell pigment signal. Pigmented fossil eggshells form a cluster (yellow) that is separated from the cluster (grey) of unpigmented fossil eggshells (Extended Data Fig. 5). The pigment fingerprint region includes—besides protoporphyrin IX and biliverdin—structurally similar PFPs as well as pigment fossilization products, which account for the distribution of samples within each cluster of fossil eggshells (pigmented and unpigmented). PC1 represents sample variation based on pigment type, concentration and PFP contents, and PC2 separates samples based on the presence or absence of eggshell pigments (Extended Data Fig. 5).



Extended Data Fig. 5 | Loadings plots for principal component axes 1–3. These loading plots are for the PCA shown in Extended Data Fig. 4. **a**, Loadings per wavelength and functional group for PC1 (57.118%). Negative loadings are associated with C–N functions, while positive loadings are associated with C–O functions. PC1 separates samples in the chemospace based on pigment type (protoporphyrin IX ratio of C–N to C=O) > biliverdin (ratio of C–N to C=O)), concentration, and

PFP contents. **b**, Loadings per wavelength and functional group for PC2 (23.841%). Negative loadings are associated with prominent PFP bands, whereas positive loadings are associated with eggshell pigments. Thus, PC2 separates samples in the chemo-space on the basis of the presence or absence of eggshell pigments. **c**, Loadings per wavelength and functional group for PC3 (9.258%).



Extended Data Fig. 6 | High-resolution Raman point measurements of eggshell sediment samples. $n = 12$ independent samples. Measurement parameters are identical to those used to process eggshell samples, to guarantee comparability ($300\text{--}2,000\text{ cm}^{-1} \pm 2\text{ cm}^{-1}$, 6 accumulations, 20 s of exposure, base-lined and normalized). Each sediment measurement was repeated three times independently and yielded similar results. The coloured dots label potential pigment-band positions (biliverdin, blue; protoporphyrin IX red). For comparison, a pigment-positive spectrum of

H. huangi eggshell (1) is provided at the bottom, followed by sediments associated with *M. peeblesorum* (YPM VP PU 023464) (2); *M. peeblesorum* (YPM PU 22523) (3); South American saltasaurid (4); French titanosaurid (5); *H. huangi* (6); Mongolian microtroodontid (7); North American troodontid (8); Chinese troodontid (9); Mongolian troodontid (10); *D. antirrhopus* (11); Mongolian enantiornithine (12); and North American ornithoid ratite (13) eggshells. None of the sediments contains substantial amounts of eggshell pigments.

Protocadherin-1 is essential for cell entry by New World hantaviruses

Rohit K. Jangra^{1,15}, Andrew S. Herbert^{2,15}, Rong Li^{3,15}, Lucas T. Jae^{4,12,15}, Lara M. Kleinfelter¹, Megan M. Slough¹, Sarah L. Barker⁵, Pablo Guardado-Calvo⁶, Gleyder Román-Sosa⁶, M. Eugenia Dieterle¹, Ana I. Kuehne², Nicolás A. Muena⁷, Ariel S. Wirchnianski^{2,13}, Elisabeth K. Nyakatura⁸, J. Maximilian Fels¹, Melinda Ng¹, Eva Mittler¹, James Pan⁵, Sushma Bharrhan¹, Anna Z. Wec^{1,14}, Jonathan R. Lai⁸, Sachdev S. Sidhu⁵, Nicole D. Tischler⁷, Félix A. Rey⁶, Jason Moffat^{5,9}, Thijn R. Brummelkamp^{4,10,11*}, Zhongde Wang^{3*}, John M. Dye^{2*} & Kartik Chandran^{1*}

The zoonotic transmission of hantaviruses from their rodent hosts to humans in North and South America is associated with a severe and frequently fatal respiratory disease, hantavirus pulmonary syndrome (HPS)^{1,2}. No specific antiviral treatments for HPS are available, and no molecular determinants of in vivo susceptibility to hantavirus infection and HPS are known. Here we identify the human asthma-associated gene protocadherin-1 (*PCDH1*)^{3–6} as an essential determinant of entry and infection in pulmonary endothelial cells by two hantaviruses that cause HPS, Andes virus (ANDV) and Sin Nombre virus (SNV). In vitro, we show that the surface glycoproteins of ANDV and SNV directly recognize the outermost extracellular repeat domain of *PCDH1*—a member of the cadherin superfamily^{7,8}—to exploit *PCDH1* for entry. In vivo, genetic ablation of *PCDH1* renders Syrian golden hamsters highly resistant to a usually lethal ANDV challenge. Targeting *PCDH1*

could provide strategies to reduce infection and disease caused by New World hantaviruses.

Hantaviruses systemically infect and replicate in endothelial cells, and the nonlytic dysregulation of these cells is thought to underlie the changes in vascular permeability that are a hallmark of the viral disease in humans^{2,9}. $\alpha_v\beta_3$ integrins have been identified as in vitro determinants of hantavirus infection¹⁰, and viral subversion of β_3 -integrin signalling in endothelial cells has been proposed to compromise vascular integrity^{9,10}. Gene-complementation experiments have yielded other receptor candidates, including β_2 integrin¹¹ and numerous components of the complement system^{12,13}. However, the roles of these host factors in animal models of HPS or in humans remain undefined. Therefore, the identities of host molecules that mediate hantavirus infection in vivo and influence pathogenesis so far remain unknown.

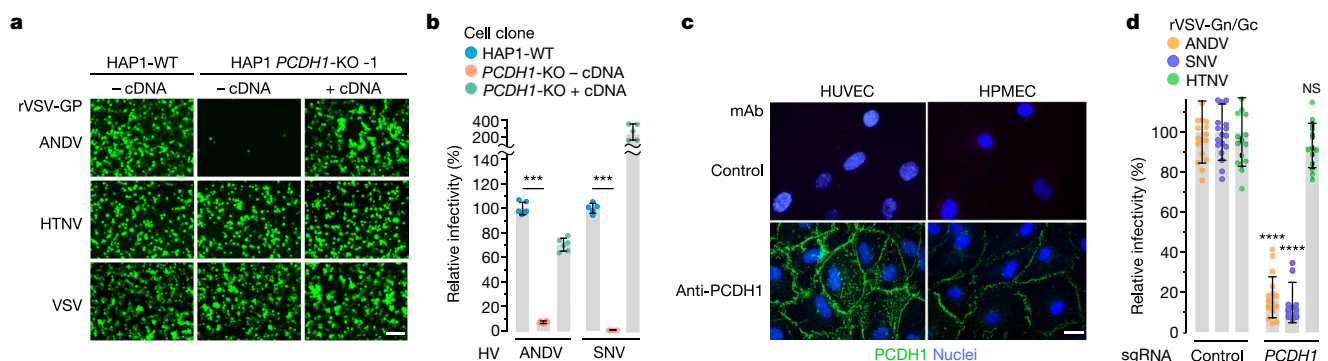


Fig. 1 | A haploid genetic screen identifies *PCDH1* as a host factor for ANDV and SNV entry and infection. **a, **b**, Relative infectivity of rVSVs bearing the indicated viral glycoproteins. Wild-type (WT) and *PCDH1*-knockout (KO) HAP1 cell lines lacking (–cDNA) or expressing (+cDNA) WT human *PCDH1* cDNA were exposed to rVSVs expressing hantavirus glycoproteins (rVSV-GPs) (**a**) or to hantaviruses (HVs) (**b**). **a**, Infected cells positive for enhanced green fluorescent protein (eGFP; pseudocoloured green) were detected by fluorescence microscopy. Representative images are shown. Scale bar, 100 μ m. **b**, Hantavirus-infected cells were detected and enumerated by immunofluorescence microscopy. Averages \pm s.d. from three experiments are shown in **b**; $n = 6$ (ANDV); $n = 5$ (SNV); WT versus *PCDH1*-KO cells, two-way ANOVA**

with Tukey's test, *** $P < 0.001$ (n indicates the number of biologically independent samples). **c**, Expression of *PCDH1* in HUVECs and HPMECs was detected by immunostaining with *PCDH1*-specific monoclonal antibody (mAb) 3305 or negative control antibody (see Extended Data Fig. 4d) and visualized by immunofluorescence microscopy. Scale bar, 20 μ m. Experiments were performed three times with similar results. **d**, HPMECs transduced to co-express the endonuclease Cas9 and control or single-guide RNAs (sgRNAs) targeting *PCDH1* were exposed to rVSVs. The results are averages \pm s.d. from five experiments; $n = 16$ for ANDV; $n = 18$ for SNV; $n = 14$ for HTNV. *PCDH1* sgRNA versus control sgRNA, two-way ANOVA with Sidak's test; NS, $P > 0.05$; **** $P < 0.0001$.

¹Department of Microbiology and Immunology, Albert Einstein College of Medicine, New York, NY, USA. ²United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD, USA. ³Department of Animal, Dairy and Veterinary Sciences, Utah State University, Logan, UT, USA. ⁴Oncode Institute, Division of Biochemistry, The Netherlands Cancer Institute, Amsterdam, The Netherlands. ⁵Donnelly Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ⁶Institut Pasteur, Structural Virology Unit and CNRS UMR3569, Paris, France. ⁷Fundación Ciencia & Vida, Laboratorio de Virología Molecular, Santiago, Chile. ⁸Department of Biochemistry, Albert Einstein College of Medicine, New York, NY, USA. ⁹Canadian Institute for Advanced Research, Toronto, Ontario, Canada. ¹⁰CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. ¹¹Cancer Genomics.nl (CGC.nl), Amsterdam, The Netherlands. ¹²Present address: Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. ¹³Present address: Department of Microbiology and Immunology and Department of Biochemistry, Albert Einstein College of Medicine, New York, NY, USA. ¹⁴Present address: Adimab LLC, Lebanon, NH, USA. ¹⁵These authors contributed equally: Rohit K. Jangra, Andrew S. Herbert, Rong Li, Lucas T. Jae. *e-mail: t.brummelkamp@nki.nl; zonda.wang@usu.edu; john.m.dye1.civ@mail.mil; kartik.chandran@einstein.yu.edu

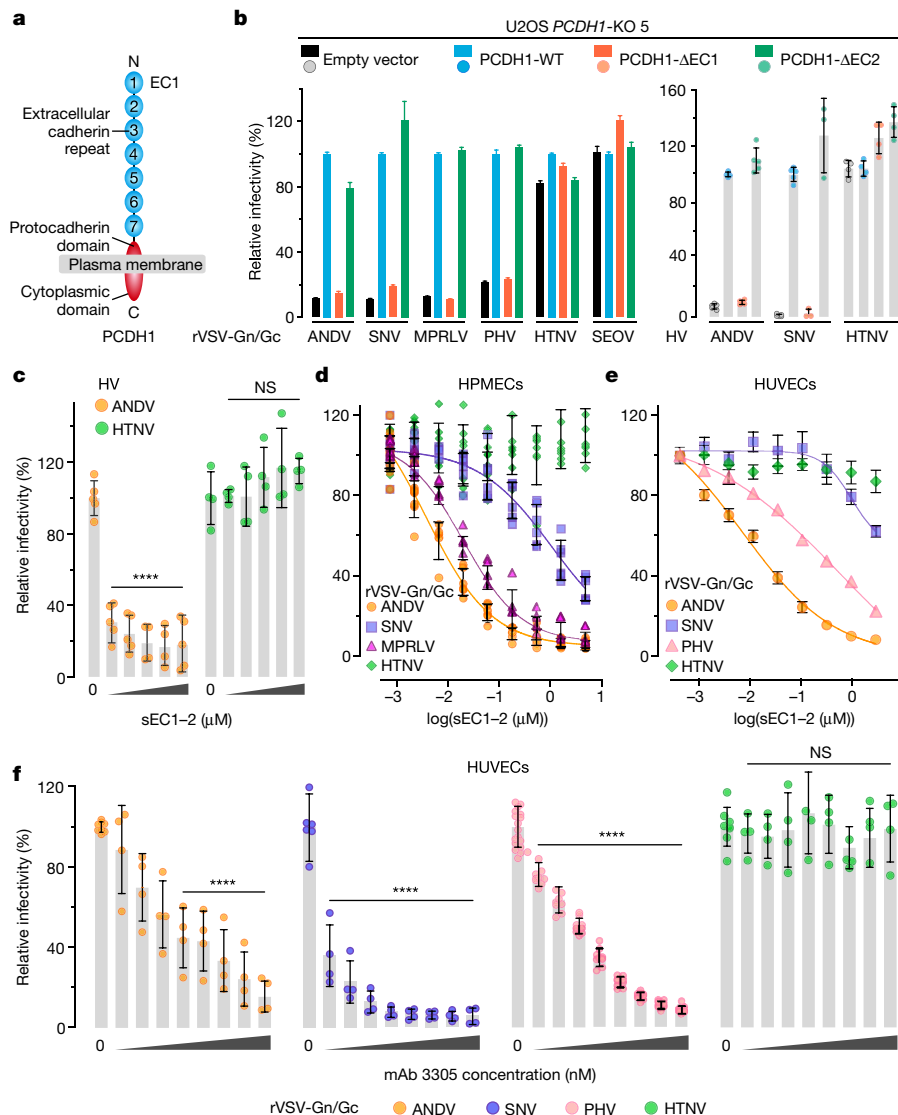


Fig. 2 | The first extracellular cadherin domain of PCDH1 is required for New World hantavirus entry and infection. **a**, Organization of PCDH1. **b**, U2OS *PCDH1*-KO cell lines complemented with the indicated PCDH1 proteins were exposed to rVSVs or hantaviruses. Left, averages \pm s.e.m.: five experiments, $n = 25$ for ANDV and HTNV; five experiments, $n = 15$ for SNV except full-length PCDH1 (four experiments, $n = 12$); four experiments, $n = 24$ for MPRLV and SEOV; three experiments, $n = 10$ for PHV. Right, averages \pm s.d.: three experiments, $n = 5$ for ANDV and SNV, except SNV Δ EC1 and Δ EC2 (two experiments, $n = 3$); two experiments, $n = 4$ for HTNV. **c**, Capacity of sEC1-2 (0–2.2 μ M) to block authentic hantavirus infection. Viruses were preincubated with sEC1-2, and then allowed to infect WT U2OS cells. Averages \pm s.d.: two experiments, $n = 4$ or 5 for ANDV; two experiments,

$n = 4$ for HTNV. Untreated versus sEC1-2-treated, two-way ANOVA with Dunnett's test; NS, $P > 0.05$; **** $P < 0.0001$. **d**, **e**, Capacity of soluble, truncated PCDH1 (sEC1-2, 0–5 μ M) to block viral entry. rVSVs were preincubated with sEC1-2, and then allowed to infect HPMECs (**d**) and HUVECs (**e**). **d**, Averages \pm s.d.: three experiments, $n = 8$ for ANDV and HTNV; $n = 6$ for SNV; $n = 4$ for MPRLV. **e**, Averages \pm s.e.m.: six experiments, $n = 15$ or 16 for ANDV and HTNV; three experiments, $n = 5$ or 6 for SNV; four experiments, $n = 8$ for PHV. **f**, Capacity of PCDH1 EC1-specific mAb 3305 to block viral entry. HUVECs were preincubated with mAb 3305 (0–680 nM), and then exposed to rVSVs. Averages \pm s.d.: three experiments, $n = 4$ for ANDV, SNV and HTNV; four experiments, $n = 9$ for PHV. Untreated versus antibody-treated by two-way ANOVA with Dunnett's test: NS, $P > 0.05$; **** $P < 0.0001$.

To systematically uncover host factors for hantavirus entry, we¹⁴ and others¹⁵ previously used a recombinant vesicular stomatitis virus bearing the ANDV Gn/Gc glycoproteins (rVSV-ANDV Gn/Gc) to perform a loss-of-function genetic screen in HAP1 haploid human cells (Extended Data Fig. 1a). These screens identified several genes involved in the sterol regulatory element binding protein (SREBP) pathway as determinants of viral entry in endothelial cells and showed that membrane cholesterol has a key role in hantavirus membrane fusion^{14,16}.

To extract hantavirus-receptor candidates from our dataset, we filtered our hits for genes that encode known plasma-membrane proteins¹⁷, and found a single gene, *PCDH1*—which encodes a cadherin-superfamily protein^{7,8}, protocadherin-1—with proposed roles in human airway function and disease^{3–6} (Extended Data Fig. 1a and

Supplementary Table 1). *PCDH1* was not a hit in any other published haploid screens for pathogen host factors^{18,19}, suggesting that it has a specific role in hantavirus entry. To evaluate this hypothesis, we used CRISPR–Cas9 genome engineering to generate cell clones deficient for PCDH1 in two human cell lines—HAP1 haploid cells (Fig. 1a, b and Extended Data Fig. 1b, c) and U2OS osteosarcoma cells (Fig. 2b and Extended Data Fig. 1d–f). *PCDH1*-knockout cells were poorly susceptible to infection by rVSVs bearing Gn/Gc glycoproteins from ANDV, SNV, Maporal virus (MPRLV) and Prospect Hill virus (PHV), all of which belong to the New World hantavirus clade. However, *PCDH1*-knockout cells remained susceptible to rVSVs bearing the VSV glycoprotein G or Gn/Gc from Hantaan virus (HTNV) and Seoul virus (SEOV)—hantaviruses in the Old World hantavirus clade that are not

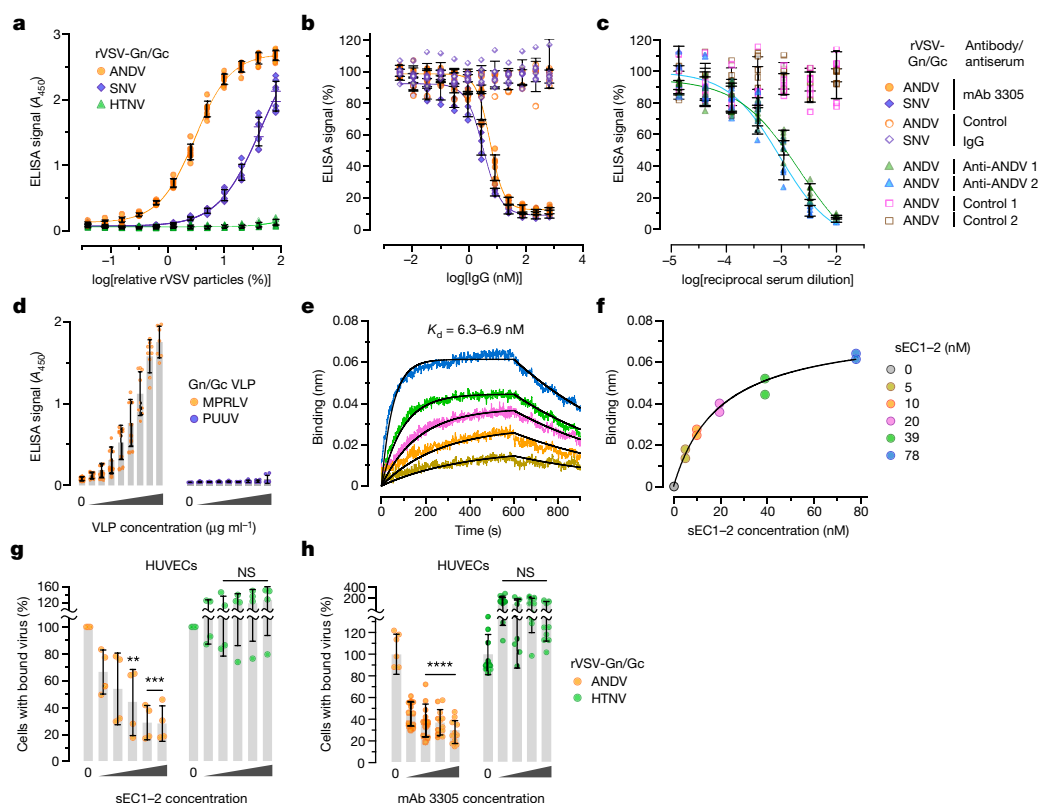


Fig. 3 | PCDH1 mediates ANDV and SNV attachment to cells by binding directly to the viral glycoproteins. **a**, Biotinylated rVSVs were added to sEC1-2-coated plates and rVSV capture was measured by ELISA. Averages \pm s.d.: four experiments, $n = 8$. **b**, The capacity of PCDH1-specific mAb 3305 to block rVSV-Gn/Gc capture by sEC1-2 was measured by ELISA as in **a**. Immobilized sEC1-2 was incubated with a control IgG or with mAb 3305 before addition of biotinylated rVSVs. Averages \pm s.d.: three experiments, $n = 6$. **c**, The capacity of Gn/Gc-reactive convalescent sera from two Chilean survivors of HPS to block binding between sEC1-2 and rVSV-Gn/Gc was measured by ELISA as in **a**. Biotinylated rVSV-ANDV Gn/Gc was incubated with serial dilutions of antisera and then added to sEC1-2-coated plates. Averages \pm s.d.: three experiments, $n = 6$. **d**, The capacity of sEC1-2 to capture purified, Strep-tagged MPRLV or PUUV VLPs ($0\text{--}170\text{ }\mu\text{g ml}^{-1}$) was measured by ELISA. Averages \pm s.d.: three experiments, $n = 7$ for PUUV, $n = 7$ or 9 for MPRLV. **e**, Sensorgrams of sEC1-2 binding to MPRLV VLPs by biolayer interferometry. Experimental curves (coloured traces) were fit using a 1/1 binding model

(black traces) to derive equilibrium dissociation constant (K_D) values. **f**, Response curve for steady-state analysis. Coloured dots correspond to the coloured curves in **e**. Results from two independent experiments are shown in **e** and **f**. **g**, Capacity of sEC1-2 to block viral attachment to cells. rVSVs bearing ANDV or HTNV Gn/Gc and labelled with functional-component spacer diacyl lipid (FSL)-fluorescein were preincubated with sEC1-2 ($0\text{--}1.6\text{ }\mu\text{M}$), and then exposed to HUVECs at 4°C for 1 hour. Cells with bound viral particles were enumerated by flow cytometry. Averages \pm s.d.: four experiments, $n = 4$. Untreated versus sEC1-2-treated, two-way ANOVA with Dunnett's test; NS, $P > 0.05$; *** $P < 0.01$; **** $P < 0.0001$. **h**, Capacity of mAb 3305 to block viral attachment to cells. HUVECs were preincubated with mAb 3305 ($0\text{--}68\text{ nM}$) at 4°C , and then exposed to DiD lipophilic-dye-labelled rVSVs bearing ANDV or HTNV Gn/Gc at 4°C . We obtained 6–12 images per coverslip; virus-bound cells were analysed with Volocity software. Averages \pm s.d.: three experiments, $n = 12$. Untreated versus antibody-treated, two-way ANOVA with Dunnett's test; NS, $P > 0.05$; **** $P < 0.0001$.

associated with HPS. The susceptibility of the *PCDH1*-knockout cell lines to Gn/Gc-dependent entry could be restored by expression of human *PCDH1* complementary DNA (Figs. 1a, b and Extended Data Fig. 1). Infections with authentic hantaviruses corroborated and extended our observations: ANDV and SNV required PCDH1 for infection, whereas HTNV did not (Figs. 1b, 2b). Thus, PCDH1 mediates Gn/Gc-dependent cell entry and infection by four New World hantaviruses—including two that are associated with HPS (ANDV and SNV)—but not by two Old World hantaviruses that are associated with haemorrhagic fever with renal syndrome.

Endothelial cells, including those of the lung microvasculature, are major targets of hantavirus infection *in vivo*^{2,20}. Consistent with this, previous work showed that PCDH1 is expressed in human airway epithelial and endothelial cells^{3,6}. Here we found abundant cell-surface expression of PCDH1 in both human primary umbilical vein endothelial cells (HUVECs) and human primary pulmonary microvascular endothelial cells (HPMECs) (Fig. 1c). Further, depletion of PCDH1 in HPMECs by CRISPR-Cas9 engineering selectively reduced infection by rVSVs that bear ANDV and SNV Gn/Gc glycoproteins (Fig. 1d), which suggests that PCDH1 has a critical role in ANDV and SNV Gn/Gc-dependent viral entry into primary endothelial cells.

PCDH1 is a type I membrane protein with seven extracellular cadherin-repeat (EC) domains and a protocadherin-specific cytoplasmic domain^{7,8,21} (Fig. 2a). To identify PCDH1 sequences that are required for hantavirus entry, we generated constructs that lack (Δ) the first or second EC domains (EC1 (human PCDH1 amino-acid residues 61–172) and EC2 (residues 173–284)) and tested their ability to complement U2OS *PCDH1*-knockout cells (Fig. 2). Although both proteins were expressed and localized to the plasma membrane (Extended Data Fig. 2), only PCDH1- Δ EC1 failed to restore infection (Fig. 2b), suggesting a role for EC1 in hantavirus entry. Reasoning that EC1 may interact with hantavirus Gn/Gc, we next expressed and purified recombinant, soluble forms of EC1–EC2 (sEC1–2) and EC1 (sEC1) (Extended Data Fig. 3a–c) and evaluated their effects on viral infection. Preincubation with sEC1–2 or sEC1 selectively inhibited entry and infection by New World hantaviruses (Fig. 2c–e and Extended Data Fig. 3d).

To directly target PCDH1 EC1, we developed and characterized a panel of monoclonal antibodies against purified PCDH1 EC domains (Extended Data Fig. 4a–c) and confirmed that they could decorate cell surfaces in a PCDH1-EC1-specific manner (Extended Data Fig. 4d). Preincubation of HUVECs (Fig. 2f) and HPMECs (Extended Data Fig. 5) with EC1-specific antibody 3305 specifically inhibited

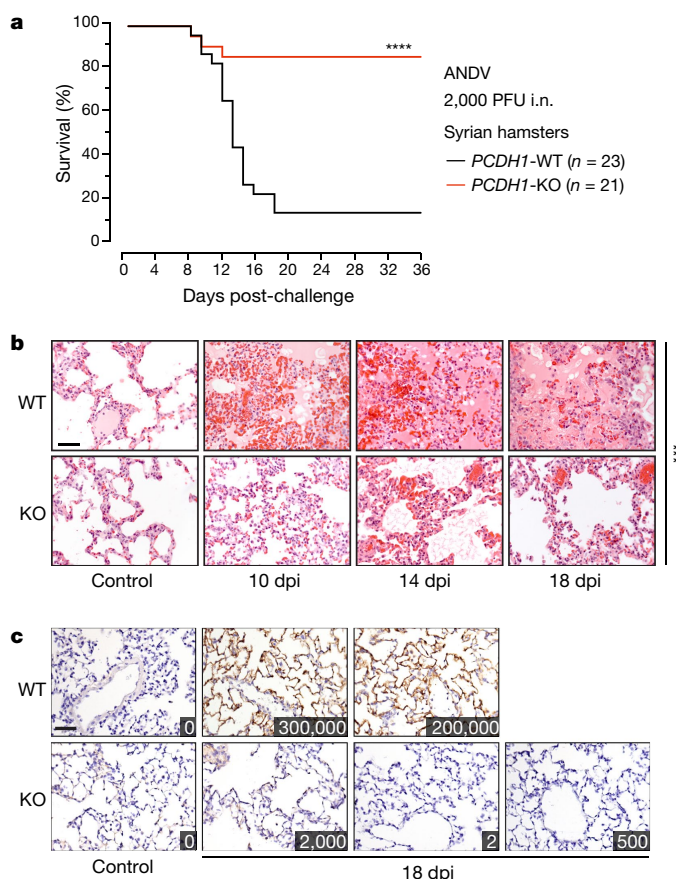


Fig. 4 | *PCDH1* is required for lethal ANDV pulmonary infection in Syrian hamsters. **a**, Wild-type hamsters (two experiments, $n = 23$) and knockout hamsters (two experiments, $n = 21$) were exposed to ANDV (target dose 2,000 plaque-forming units (PFU); actual dose 1,500 PFU) via the intranasal (i.n.) route, and mortality was monitored for 36 days. *PCDH1*-KO versus *PCDH1*-WT, two-sided Mantel-Cox test; **** $P < 0.0001$. **b**, **c**, Representative images of lung sections from control and ANDV-challenged Syrian hamsters were stained with haematoxylin and eosin (**b**) or with a mAb targeting the ANDV nucleoprotein (**c**). Insets in **c** show viral loads in lung tissue (in units of viral genome equivalents per μg of total RNA). Experiments were performed twice with similar results. Scale bar, 100 μm (**b**, **c**). WT versus KO ($n = 3$) at 18 days post-infection (dpi), two-way ANOVA with Dunnett's test; **** $P < 0.001$.

Gn/Gc-dependent entry by New World viruses. Taken together, these findings suggest that an interaction between viral Gn/Gc glycoproteins and the first extracellular cadherin-repeat domain of *PCDH1*, EC1, drives entry by four New World hantaviruses.

We used four distinct approaches to evaluate the hypothesis that New World hantavirus Gn/Gc glycoproteins directly recognize and engage *PCDH1* EC1. First, we infected cells with rVSVs bearing ANDV and HTNV Gn/Gc and incubated the resulting Gn/Gc-displaying cells with sEC1-2 (Extended Data Figs. 6, 7a, b). Only ANDV Gn/Gc-expressing cells could be decorated with sEC1-2, and sEC1-2 capture was sensitive to the EC1-specific antibody 3305 (Extended Data Fig. 7c). Second, we incubated sEC1-2 immobilized on ELISA plates with pre-titrated amounts of viral particles bearing Gn/Gc. sEC1-2 could selectively capture ANDV and SNV rVSV Gn/Gc from solution (Fig. 3a), in a manner that was sensitive to both antibody 3305 (Fig. 3b) and convalescent sera from two Chilean survivors of HPS²² (Fig. 3c), providing evidence that the interaction is specific to both *PCDH1* and Gn/Gc. Third, we incubated biotinylated rVSV Gn/Gc particles with purified sEC1-2 and recovered the particles with streptavidin beads. rVSV-ANDV Gn/Gc co-precipitated substantially higher levels of sEC1-2 compared with the negative control (beads alone), whereas rVSV-HTNV Gn/Gc did not (Extended Data Fig. 7d). Fourth, purified, recombinant virus-like

particles (VLPs)²³ constituted with Gn/Gc from MPRLV (Extended Data Fig. 8a, b)—but not the Old World hantavirus Puumala virus (PUUV)—could recognize sEC1-2 in the capture ELISA described above (Fig. 3d), and MPRLV VLPs bound to sEC1-2 with an equilibrium dissociation constant (K_d) of around 7 nM, as determined by biolayer interferometry (BLI) (Fig. 3e, f and Extended Data Fig. 8c). Finally, we assessed the capacity of the interaction-blocking reagents sEC1-2 and antibody 3305 to interfere with viral attachment to the cell surface (Fig. 3g, h). Both sEC1-2 (Fig. 3g) and the antibody (Fig. 3h) specifically blocked attachment of rVSV-ANDV Gn/Gc particles to HUVECs. Together, these findings provide evidence that New World hantavirus glycoproteins directly recognize the EC1 domain of *PCDH1* with nanomolar affinity and exploit this interaction to mediate hantavirus attachment to the cell surface.

Finally, we sought to evaluate the importance of *PCDH1* to hantavirus infection and disease pathogenesis in the Syrian golden hamster (*Mesocricetus auratus*)—the gold-standard rodent model for lethal HPS²⁴. The human and Syrian hamster *PCDH1* orthologues are highly conserved in amino-acid sequence (with more than 97% sequence identity in EC1), suggesting that interactions of *PCDH1* with both hantavirus Gn/Gc and antibody 3305 are likely to be preserved in hamster cells. Indeed, sEC1-2 and sEC1—but not sEC2—selectively blocked rVSV-ANDV Gn/Gc infection of hamster primary lung endothelial cells, as did the EC1-specific monoclonal antibody, which confirms that Gn/Gc-*PCDH1* recognition drives ANDV entry into endothelial cells from this species (Extended Data Fig. 9a, b).

Next, we used CRISPR-Cas9 genome engineering to generate hamsters carrying a germline-transmitted, single-nucleotide deletion in *PCDH1* (Extended Data Fig. 10a–c). This frameshift mutation causes multiple premature stop codons in *PCDH1* and its genetic inactivation (knockout), as confirmed by immunoblotting of isolated lung tissue from animals homozygous for the knockout allele (Extended Data Fig. 10d). *PCDH1*-knockout hamsters were viable and fertile, and morphological and histopathological analysis revealed no lesions in the lungs (Fig. 4c, left) or in other sampled organs, suggesting that *PCDH1* loss is not associated with tissue malformation or injury. We then subjected wild-type and *PCDH1*-knockout hamsters to intranasal challenge with ANDV (Fig. 4a). Wild-type hamsters largely succumbed to ANDV infection, as previously shown^{25,26}. By contrast, *PCDH1*-knockout hamsters largely survived ANDV challenge, indicating that the loss of *PCDH1* is highly protective (Fig. 4a). Measurement of viral titres in sera collected from challenged animals on day 14—near the onset of lethality—indicated reduced levels of viraemia in knockout animals (Extended Data Fig. 10e). Histopathological analysis of lungs isolated from challenged wild-type animals (Fig. 4b) revealed severe interstitial pneumonitis, perivascular inflammation associated with macrophage and neutrophil infiltration, oedema, haemorrhage, and fibrin deposition. By contrast, knockout animals had only mild inflammation with thickening of alveolar septae and occasional macrophage and neutrophil infiltration (Fig. 4b). Furthermore, lung tissue collected from wild-type but not knockout animals at 18 days post-infection showed extensive staining for viral antigen in endothelial and alveolar septal cells (Fig. 4c). Concordantly, we measured substantially higher levels of viral RNA in lung homogenates from wild-type animals relative to their knockout counterparts (Fig. 4c). Blocking the hantavirus-*PCDH1* interaction both in vitro and in vivo did not fully prevent viral attachment, infection and histopathology, which suggests the existence of *PCDH1*-independent entry pathways. Nevertheless, our findings establish a role for *PCDH1* in ANDV infection in vivo and in the development of disease in the Syrian hamster model.

Herein we have used global gene disruption in human cells to discover a protein from the cadherin superfamily, *PCDH1*, as a critical host factor for cell entry and infection by multiple New World hantaviruses, including ANDV and SNV—the major aetiological agents of HPS in the Americas. Pharmacological disruption of the interaction between the viral glycoproteins and *PCDH1* might afford the development of anti-HPS therapeutics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0702-1>.

Received: 22 February 2018; Accepted: 20 September 2018;

Published online 21 November 2018.

- MacNeil, A., Nichol, S. T. & Spiropoulou, C. F. Hantavirus pulmonary syndrome. *Virus Res.* **162**, 138–147 (2011).
- Zaki, S. R. et al. Hantavirus pulmonary syndrome. Pathogenesis of an emerging infectious disease. *Am. J. Pathol.* **146**, 552–579 (1995).
- Kozu, Y. et al. Protocadherin-1 is a glucocorticoid-responsive critical regulator of airway epithelial barrier function. *BMC Pulm. Med.* **15**, 80 (2015).
- Mortensen, L. J., Kreiner-Møller, E., Hakonarson, H., Bønnelykke, K. & Bisgaard, H. The PCDH1 gene and asthma in early childhood. *Eur. Respir. J.* **43**, 792–800 (2014).
- Toncheva, A. A. et al. Genetic variants in protocadherin-1, bronchial hyper-responsiveness, and asthma subphenotypes in German children. *Pediatr. Allergy Immunol.* **23**, 636–641 (2012).
- Koppelman, G. H. et al. Identification of *PCDH1* as a novel susceptibility gene for bronchial hyperresponsiveness. *Am. J. Respir. Crit. Care Med.* **180**, 929–935 (2009).
- Gul, I. S., Hulpiau, P., Saeys, Y. & van Roy, F. Evolution and diversity of cadherins and catenins. *Exp. Cell Res.* **358**, 3–9 (2017).
- Sotomayor, M., Gaudet, R. & Corey, D. P. Sorting out a promiscuous superfamily: towards cadherin connectomics. *Trends Cell Biol.* **24**, 524–536 (2014).
- Mackow, E. R. & Gavrilovskaya, I. N. Hantavirus regulation of endothelial cell functions. *Thromb. Haemost.* **102**, 1030–1041 (2009).
- Gavrilovskaya, I. N., Shepley, M., Shaw, R., Ginsberg, M. H. & Mackow, E. R. $\beta 3$ integrins mediate the cellular entry of hantaviruses that cause respiratory failure. *Proc. Natl Acad. Sci. USA* **95**, 7074–7079 (1998).
- Rafferty, M. J. et al. $\beta 2$ integrin mediates hantavirus-induced release of neutrophil extracellular traps. *J. Exp. Med.* **211**, 1485–1497 (2014).
- Buranda, T. et al. Recognition of decay accelerating factor and $\alpha v \beta 3$ by inactivated hantaviruses: toward the development of high-throughput screening flow cytometry assays. *Anal. Biochem.* **402**, 151–160 (2010).
- Krautkrämer, E. & Zeier, M. Hantavirus causing hemorrhagic fever with renal syndrome enters from the apical surface and requires decay-accelerating factor (DAF/CD55). *J. Virol.* **82**, 4257–4264 (2008).
- Kleinfelter, L. M. et al. Haploid genetic screen reveals a profound and direct dependence on cholesterol for hantavirus membrane fusion. *MBio* **6**, e00801-15 (2015).
- Petersen, J. et al. The major cellular sterol regulatory pathway is required for Andes virus infection. *PLoS Pathog.* **10**, e1003911 (2014).
- Tischler, N. D., Gonzalez, A., Perez-Acle, T., Roseblatt, M. & Valenzuela, P. D. Hantavirus Gc glycoprotein: evidence for a class II fusion protein. *J. Gen. Virol.* **86**, 2937–2947 (2005).
- Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
- Pillay, S. & Carette, J. E. Hunting viral receptors using haploid cells. *Annu. Rev. Virol.* **2**, 219–239 (2015).
- Staring, J., Raaben, M. & Brummelkamp, T. R. Viral escape from endosomes and host detection at a glance. *J. Cell Sci.* **131**, jcs216259 (2018).
- Nolte, K. B. et al. Hantavirus pulmonary syndrome in the United States: a pathological description of a disease caused by a new agent. *Hum. Pathol.* **26**, 110–120 (1995).
- Sano, K. et al. Protocadherins: a large family of cadherin-related molecules in central nervous system. *EMBO J.* **12**, 2249–2256 (1993).
- Tischler, N. D., Galeno, H., Roseblatt, M. & Valenzuela, P. D. Human and rodent humoral immune responses to Andes virus structural proteins. *Virology* **334**, 319–326 (2005).
- Acuña, R. et al. Hantavirus Gn and Gc glycoproteins self-assemble into virus-like particles. *J. Virol.* **88**, 2344–2348 (2014).
- Safronetz, D., Ebihara, H., Feldmann, H. & Hooper, J. W. The Syrian hamster model of hantavirus pulmonary syndrome. *Antiviral Res.* **95**, 282–292 (2012).
- Hooper, J. W., Ferro, A. M. & Wahl-Jensen, V. Immune serum produced by DNA vaccination protects hamsters against lethal respiratory challenge with Andes virus. *J. Virol.* **82**, 1332–1338 (2008).
- Safronetz, D. et al. Pathogenesis and host response in Syrian hamsters following intranasal infection with Andes virus. *PLoS Pathog.* **7**, e1002426 (2011).

Acknowledgements See Supplementary Information for grants supporting this work. We thank A. Beck and the Einstein Histopathology Core for histopathology support; S. Garforth and the Einstein Macromolecular Therapeutics Development Facility for assistance with SEC-MALS; G. Pehau-Arnaudet, Institut Pasteur Paris, for electron microscopy of MPRLV VLPs; H. Galeno, Instituto de Salud Pública de Chile, for convalescent sera from Chilean patients with hantavirus; and M. Evans for feedback and discussions. Opinions, conclusions, interpretations and recommendations are those of the authors and are not necessarily endorsed by the US Army. The mention of trade names or commercial products does not constitute endorsement or recommendation for use by Department of the Army or Department of Defense.

Reviewer information Nature thanks B. Lee, G. Schönrich and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions R.K.J., L.T.J., K.C., T.R.B., A.S.H. and J.M.D. conceived the study. R.K.J., M.M.S. and K.C. generated rVSVs expressing hantavirus Gn/Gc proteins. L.T.J. and T.R.B. performed the haploid genetic screen and identified hits. R.K.J. genetically validated PCDH1 as a hantavirus entry factor with assistance from M.E.D., and performed biosafety level 2 virologic and mechanistic studies with assistance from M.M.S. In vitro studies with authentic hantaviruses were performed by A.S.H., A.I.K., A.S.W. and J.M.D. R.K.J. generated PCDH1 variants with assistance from J.M.F. and M.M.S. R.K.J., M.M.S. and M.E.D. carried out PCDH1–Gn/Gc binding studies. G.R.-S developed the affinity purification of MPRLV and PUUV VLPs. P.G.-C., G.R.-S., F.A.R., E.K.N. and J.R.L. performed biolayer interferometry-based binding studies. S.L.B., J.P., J.M. and S.S.S. generated and epitope-mapped PCDH1-specific monoclonal antibodies by phage display, and S.L.B., R.K.J. and A.Z.W. assisted in expression and characterization of monoclonal antibodies. L.M.K. performed studies to assess the subcellular distribution of PCDH1 variants and to determine the cell-biological functions of PCDH1, with assistance from R.K.J. and E.M. N.A.M. and N.D.T. developed and characterized Gn/Gc-specific monoclonal antibodies. R.K.J. designed and validated hamster-specific guide RNAs with assistance from M.N. R.L. and Z.W. generated and bred *PCDH1*-knockout hamsters, and R.L. and R.K.J. genotyped them. A.S.H., A.I.K. and J.M.D. performed hamster challenge studies. S.B. assisted in hamster tissue processing and initial optimizations of immunohistochemistry staining. K.C. and R.K.J. wrote the paper with contributions from all authors.

Competing interests R.K.J., L.T.J., K.C. and T.R.B. are named inventors on US patent application US20170173141A1, covering methods to treat hantavirus infections by targeting PCDH1 (the entry factor described here), the application being assigned to Albert Einstein College of Medicine. T.R.B. is co-founder and Scientific Advisory Board member of Haplogen GmbH, and co-founder and managing director of Scenic Biotech BV. F.A.R. is a consultant for Flagship Pioneering.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0702-1>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0702-1>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to T.R.B. or Z.W. or J.M.D. or K.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Cells. Human U2OS osteosarcoma cells, 293T human embryonic kidney fibroblast cells and Freestyle-293F suspension-adapted HEK-293 cells were obtained from ATCC. U2OS and 293T lines were cultured in modified McCoy's 5A medium and high-glucose Dulbecco's modified Eagle medium (DMEM), respectively (Thermo Fisher), supplemented with 10% fetal bovine serum (FBS; Atlanta Biologicals), 1% GlutaMAX (Thermo Fisher) and 1% penicillin-streptomycin (Thermo Fisher). HAP1 cells were cultured in Iscove's modified Dulbecco's medium (IMDM; Thermo Fisher), supplemented as above. HUVECs (Lonza) were cultured in endothelial growth medium (EGM) supplemented with EGM-SingleQuots (Lonza). HPMECs (Promocell) were cultured in MV2 endothelial cell growth medium (Promocell). Syrian hamster primary lung microvascular endothelial cells (Cell Biologicals) were cultured in complete endothelial cell medium (Cell Biologicals). All cell lines were authenticated by their respective manufacturers or vendors, and no additional authentication was performed. Cell lines from ATCC were authenticated by human short-tandem-repeat analysis. All cell lines were routinely screened for mycoplasma contamination (once monthly) and found to be negative. Adherent cell lines were maintained in a humidified 37°C, 5% CO₂ incubator. Shake flasks of 293F cells were cultured in Gibco FreeStyle 293 expression medium (Thermo Fisher) at 37°C and 8% CO₂.

Generation of PCDH1-knockout cell lines. A CRISPR sgRNA was designed to target the *PCDH1* gene (NCBI Gene identification number 5,097), nucleotides 9,776–9,798 (target sequence 5'-GTTTGTAGCGGCCCTCCTATGAGG-3', with the protospacer acceptor motif (PAM) sequence underlined), in exon 2 of the *PCDH1* isoform 1 precursor messenger RNA (GenBank accession number NM_002587). This sgRNA sequence was cloned into Addgene vector 41,824 (gRNA cloning vector; a gift from G. Church²⁷) and used in conjunction with a Cas9-encoding plasmid (Addgene plasmid 41,815; a gift from G. Church²⁷, Harvard Medical School and Massachusetts Institute of Technology) to generate HAP1 and U2OS *PCDH1*-KO single-cell clones, as described¹⁴. The *PCDH1* genotypes of these cell lines were verified by Sanger sequencing of cloned polymerase chain reaction (PCR) amplicons flanking the sgRNA target sequence, as described¹⁴, except that the *PCDH1*-specific primers 5'-CCTTAGGGGTACAGGAAAC-3' and 5'-GACAACACACCAACTTCGC-3' were used for PCR amplification of genomic DNA. HAP1 and U2OS *PCDH1*-KO clones both contained a single nucleotide insertion (T) after *PCDH1* nucleotide 9,792 (*PCDH1* amino-acid residue Y285). The resulting frame-shifted transcript is predicted to encode a truncated 285-amino-acid polypeptide.

CRISPR-Cas9 engineering of HPMECs. The human *PCDH1*-specific sgRNA sequence mentioned above was cloned into Addgene vector 52,961 (lentiCRISPR v2; a gift from F. Zhang²⁸, Massachusetts Institute of Technology and Broad Institute) and was used to produce lentiviruses following triple transfection of Addgene vector 12,260 (psPAX2; a gift from D. Trono, École Polytechnique Fédérale de Lausanne) and a plasmid expressing VSV glycoprotein G in 293FT cells. The empty vector was used as a negative control. Lentiviral supernatants were filtered and applied to monolayers of HPMECs. At 5–7 days post-transduction, cells were exposed to rVSVs bearing hantavirus Gn/Gc, and infection was monitored by fluorescence microscopy.

rVSVs and infections. rVSVs expressing eGFP and bearing Gn/Gc from ANDV¹⁴, SNV¹⁴, HTNV¹⁴, SEOV (GenBank accession number M34882.1) and PHV (GenBank accession number X55129.1) were engineered, rescued and propagated as described^{14,29}. Cells were exposed to rVSVs at a multiplicity of infection (MOI) of 0.02–0.2 infectious units (IU) per cell (see below for specific MOIs and times post-infection used in each experiment), and viral infectivity was measured by automated enumeration of eGFP-positive cells from captured images using a custom analysis pipeline in CellProfiler³⁰ or directly from multiwell plates using a CellInsight CX5 automated fluorescence microscope and onboard HCS Studio software (Thermo Fisher). Throughout this paper, MOIs for rVSVs are based on the calculations for the particular cell lines on which the experiments were performed.

For Figs. 1, 2, infection MOIs, times post-infection at which viral infectivity was measured, and the percentage of infected cells that corresponds to 100% relative infectivity are as follows. Values of these parameters for the Extended Data Figs. are presented in each figure legend. Figure 1a: cells were infected with rVSVs at a MOI of 0.02 IU per cell and scored for infection at 20 hours post-infection (hpi). Figure 1d: cells were infected with rVSVs at a MOI of 0.2 IU per cell and scored for infection at 9 hpi; 100% relative infectivity corresponds to 10–20% infected cells. Figure 2b (left): cells were exposed to rVSVs at a MOI of 0.02 IU per cell and scored for infection at 20–24 h; 100% relative infectivity corresponds to 20–40% infected cells for rVSVs (but 8–12% for rVSV-PHV). Figure 2d–f: cells were exposed to rVSVs at a MOI of 0.2 IU per cell and scored for infection at 9 hpi; 100% relative infectivity corresponds to 15–20% infected cells.

For soluble PCDH1 and PCDH1-specific-antibody experiments, pretitrated amounts of rVSV particles (0.1–0.2 IU per cell; see above and the next section for

specific MOIs for rVSVs and authentic hantaviruses, respectively) were incubated with increasing concentrations of test reagent at room temperature or 37°C, respectively, for 1 h before addition to cell monolayers in 96-well plates. Viral infectivities were measured as above.

Authentic hantaviruses and infections. ANDV strain Chile-9717869, SNV strain CC107 and HTNV strain 76-118 were propagated in Vero E6 cells as described^{31–33}. Hantavirus infections were performed, and infected cells were immunostained for viral antigen, as described¹⁴. In brief, HAP1 and U2OS cells were exposed to virus at a MOI of 0.5–3 PFU per cell (ANDV and HTNV) or fluorescent focus-forming units (FFU) per cell (SNV) (see below for specific MOIs), and viral infectivity was determined by immunostaining of formalin-fixed cells at 72 h post-infection (see below for specific times) with rabbit polyclonal sera specific for ANDV, HTNV or SNV nucleoproteins (NR-9673, NR-9674 and NR-12152 (all from BEI Resources), respectively). Images were acquired at 20 fields per well with a ×20 objective on an Operetta high-content imaging device (PerkinElmer). Images were analysed with a customized scheme built from image-analysis functions present in Harmony software, and the percentage of infected cells was determined using the analysis functions. MOI values for authentic hantaviruses are based on the titres estimated from infections of Vero cell monolayers.

Infection MOIs, times post-infection at which viral infectivity was measured, and the percentage of infected cells that corresponds to 100% relative infectivity in each figure panel are as follows. Figure 1b: cells were exposed to authentic hantaviruses at a MOI of 1 PFU (ANDV and HTNV) or 1 FFU (SNV) per cell and scored for infection at 72 hpi; 100% relative infectivity corresponds to 10–20% infected cells. Figure 2b (right): cells were exposed to authentic hantaviruses at a MOI of 1 PFU per cell (ANDV and HTNV) or 3 FFU per cell (SNV) and scored for infection at 72 hpi; 100% relative infectivity corresponds to 10–20% infected cells. Figure 2c, cells were exposed to authentic hantaviruses at a MOI of 1 PFU per cell; cells were scored for infection at 72 hpi; 100% relative infectivity corresponds to 10–20% infected cells.

Truncated and soluble PCDH1 variants. A cDNA encoding human PCDH1 (isoform 1, GenBank accession number NM_002587) was synthesized in frame with Myc and Flag epitope tags at the carboxyl terminus (Epoch Biolabs) and cloned into the pBABE-puro retroviral vector³⁴. PCDH1 constructs engineered to lack the first or second extracellular cadherin repeats (Δ EC1, Δ 61–172 amino-acid residues; Δ EC2, Δ 173–284 residues) were also cloned into this vector. Constructs encoding soluble (secreted) PCDH1 variants (GenBank accession number NM_002587) were generated by cloning the following sequences into the pcDNA3.1 mammalian expression vector (Thermo Fisher): EC1 (residues 1–172), EC2 (residues 1–60, 172–284), and EC1–EC2 (residues 1–284), each in frame with a C-terminal GSG linker, followed by Myc, Flag and decahistidine tags (Extended Data Fig. 3a, b), or with a GSG linker followed by a decahistidine tag alone (Extended Data Fig. 3c). Each construct also retained the endogenous PCDH1 amino-terminal signal sequence (residues 1–60). Sequences of all the plasmids were verified by Sanger sequencing.

Stable cell populations expressing PCDH1 variants. HAP1 and U2OS *PCDH1*-KO cells ectopically expressing the above PCDH1 variants were generated by transduction with pBABE-puro vectors. Retroviruses packaging the transgenes were produced by triple transfection in 293T cells¹⁴, and target cells were directly exposed to sterile-filtered (0.45 μ m) retrovirus-laden supernatants in the presence of polybrene (6 μ g ml⁻¹). Transduced cell populations were selected with puromycin (2 μ g ml⁻¹), and transgene expression was confirmed by immunostaining.

Detection and measurement of cell-surface PCDH1 expression. U2OS cells were seeded on six-well plates, chilled on ice and incubated for 1 h with PCDH1 EC7-specific monoclonal antibody 3677 or human IgG (negative control) diluted to 5 μ g ml⁻¹ in cold U2OS medium containing 1% HEPES. Cells were washed extensively with cold Hank's balanced salt solution (HBSS), scraped off the plate, fixed with 4% paraformaldehyde and stained with anti-human IgG1 conjugated to Alexa Fluor-488 for 1 h. Cells were washed with phosphate-buffered saline (PBS) and analysed by flow cytometry.

Expression and purification of secreted PCDH1 variants. Secreted PCDH1 variants cloned into pcDNA3.1 (see above) were expressed in 293F cells in shake flasks by transient transfection with linear polyethyleneimine (Polysciences), as described³⁵, and purified by nickel-chelation chromatography. In brief, clarified cell supernatants were stirred overnight at 4°C with nickel-NTA resin (2–3 ml packed resin per 600 ml supernatant). Beads were then retrieved, washed with PBS containing 50 mM imidazole, and eluted with PBS containing 250 mM imidazole. The eluted protein was buffer-exchanged with PBS, concentrated, and stored in aliquots at –20°C. Purity of the secreted PCDH1 variants was determined by size-exclusion chromatography with multi-angle light scattering (SEC-MALS) and SDS-PAGE, stained with Bio-Safe Coomassie G-250 Stain (BioRad) and imaged on a ChemiDoc Touch Imaging System (BioRad).

Isolation and characterization of anti-PCDH1 antibodies. The panel of PCDH1-specific mAbs was isolated from a phage-displayed synthetic human

antigen-binding fragment (Fab) library (Library F)³⁶. Binding selections, phage ELISAs and Fab protein purification were performed as described³⁷. In brief, phage particles displaying the Fabs from Library F were cycled through rounds of binding selections with purified PCDH1–Fc fusion proteins (either full-length extracellular domain (EC1–7; residues 58–851) or EC2–7 (residues 190–851)) immobilized on 96-well Maxisorp Immunoplates (Thermo Fisher) (Extended Data Fig. 4). After four rounds of selection, phages were produced from individual clones grown in a 96-well format and phage ELISAs were performed to detect specific binding clones. Clones with positive binding were subjected to DNA sequencing. The DNAs encoding for variable heavy- and light-chain domains of the positive binders were cloned into vectors designed for the production of Fabs or light chain (human Kappa) or heavy chain (human IgG1). Fabs were expressed from bacterial cells and IgGs from Expi293F cells (Thermo Fisher). Fab and IgG proteins were affinity-purified on protein A affinity columns (GE Healthcare) as described³⁷.

Generation of anti-Gn/Gc monoclonal antibody. To detect native Gn/Gc in virus particles or at the cell surface, hybridoma cell clones were prepared from splenocytes of Balb/c mice (12 weeks old) following four immunizations with 20 µg of ANDV virus-like particles (VLPs)²³ in complete and incomplete Freund's adjuvant (Sigma-Aldrich) using standard techniques. Clones were screened for reactivity to ANDV VLPs by ELISA and positive clones were further subcloned. 293FT cells (Thermo Fisher) expressing ANDV³⁸ or HTNV³⁹ Gn/Gc were used to select and characterize the 1E11/D3 hybridoma clone by flow cytometry using an AlexaFluor-488 conjugated goat anti-mouse secondary antibody (Thermo Fisher). For total protein expression, cells were fixed and permeabilized with Triton X-100 before staining with the antibody. For analysis of cell-surface expression, cells were stained with primary antibody before fixing and no permeabilization step was involved (Extended Data Fig. 6).

Cell-based assays for rVSV-Gn/Gc–PCDH1 binding. To visualize the binding of sEC1–2 to cells expressing hantavirus Gn/Gc, we infected U2OS cells with rVSVs bearing ANDV or HTNV Gn/Gc. At 12–14 h post-infection, cells were washed with cold PBS, and blocked with cold 10% FBS in PBS at 4°C for 30 min. Cells were then stained for surface expression of viral Gn/Gc with mouse mAb anti-hantavirus Gn/Gc 1E11/D3 followed by anti-mouse Alexa Fluor 568 (Thermo Fisher) for 1 h each at 4°C. In parallel, rVSV-ANDV/HTNV Gn/Gc-infected cells were incubated with purified Flag-tagged sEC1–2 (5 µM for 1 h at 4°C, and then washed extensively with PBS. To visualize bound sEC1–2, we fixed cells with 4% formaldehyde (Sigma-Aldrich), and stained them with anti-Flag M2 mouse mAb (Sigma-Aldrich) and secondary antibody anti-mouse Alexa Fluor 568 (Thermo Fisher) for 1 h each at 4°C. To test the ability of mAb 3305 to block sEC1–2 binding to hantavirus Gn/Gc expressing cells, we preincubated 200 nM sEC1–2 with either control human IgG or mAb 3305 (5–500 nM) for 1 h at 4°C and performed immunostaining for sEC1–2 as described above. Cells were visualized by fluorescence microscopy.

Immunofluorescence microscopy and co-immunostaining. For visualization of PCDH1, U2OS or primary endothelial cells plated on gelatin or fibronectin-coated glass coverslips were chilled on ice for 5 min and stained with control (hlgG ctrl) or PCDH1-specific mAb (5 µg ml^{−1}) for 1 h at 4°C in cell-growth medium containing 10% FBS. Cells were then washed with cold PBS, fixed with 4% paraformaldehyde for 5 min on ice, and then stained with anti-human IgG secondary antibodies conjugated to Alexa Fluor-555 or Alexa Fluor-488 (Thermo Fisher) for 1 h each at room temperature. For immunostaining with an anti-Flag antibody, paraformaldehyde-fixed cells were permeabilized with 0.1% Triton X-100 in PBS for 5 min, and blocked with 10% FBS in PBS at room temperature for 30 min. Cells were then stained with anti-Flag M2 mouse mAb (Sigma-Aldrich) and an anti-mouse IgG secondary antibody conjugated to Alexa Fluor-568 (Thermo Fisher), for 1 h each at 4°C. Coverslips were mounted in Prolong with DAPI (Thermo Fisher), and cells were imaged with a Zeiss Axio Observer inverted microscope under a ×63 objective.

rVSV-Gn/Gc–PCDH1 binding and competition ELISAs. The capacity of rVSV-Gn/Gc to recognize sEC1–2 was determined by capture ELISA (Fig. 3a). In brief, high-protein-binding 96-well ELISA plates (Corning) were coated with purified sEC1–2 (100 ng per well) overnight at 4°C and blocked with 5% nonfat dry milk in PBS (PBS-milk). The membranes of rVSV particles normalized for protein concentration were labelled with a short-chain phospholipid probe, functional-component spacer diacyl lipid conjugated to biotin (FSL–biotin; Sigma-Aldrich), as described¹⁴, and then pretitrated for biotin content by streptavidin ELISA. The sEC1–2-coated plates were incubated with dilutions of biotinylated viral particles in PBS-milk for 1 h at 37°C. Bound rVSVs were detected by incubation with a streptavidin–horseradish peroxidase (HRP) conjugate.

The capacity of PCDH1-specific mAb 3305 to inhibit virus–PCDH1 binding in the capture ELISA was determined as above, except that, first, the sEC1–2 coated plate was incubated with the indicated dilutions of each mAb before addition of viral particles; and second, a single concentration of viral particles, normalized for binding to the sEC1–2-coated plate, was then added to the plates.

The capacities of ANDV Gn/Gc-reactive convalescent sera (serial numbers 703,328 and 703,329)²² and two ANDV-seronegative human serum controls to inhibit virus–PCDH1 binding in the capture ELISA was determined as above, except that rVSV-Gn/Gc particles at a single concentration normalized for binding to sEC1–2 were incubated with the indicated dilutions of each serum before their addition to sEC1–2-coated plates. Human sera were provided by H. Galeno, Instituto de Salud Pública de Chile.

Co-precipitation assay for rVSV-Gn/Gc–PCDH1 association. The capacity of rVSV-Gn/Gc to recognize Flag-tagged sEC1–2 was determined by co-precipitation. FSL-biotin-labelled rVSV particles bearing ANDV or HTNV Gn/Gc were immobilized on streptavidin-coated magnetic beads (Spherotech) by incubation for 1 h at room temperature. Coated beads were then blocked with 5% nonfat dry milk in PBS for 30 min and washed twice with PBS. Beads were incubated with purified sEC1–2 (37 µg ml^{−1}) for 1 h at room temp and washed five times with PBS. Captured proteins were eluted by heating the beads at 98°C for 5 min with Laemmli sample buffer, subjected to SDS–PAGE and immunoblotting using an anti-Flag M2 mAb–HRP conjugate (Sigma-Aldrich) for sEC1–2 detection and mAb 23H12⁴⁰, followed by an anti-mouse IgG secondary antibody–HRP conjugate (Santa Cruz) to detect the matrix protein M in rVSV particles.

Preparation and characterization of MPRLV and PUUV VLPs. The DNA fragments encoding MPRLV Gn/Gc (residues 21 to 1138) and PUUV Gn/Gc (residues 20 to 1148) were amplified by PCR and fused in-frame to a double streptavidin tag (at the N-terminus of Gn). HEK293 cell lines expressing the VLPs were established after selection of individual clones in the presence of geneticin at 0.5 mg ml^{−1}. The cells were then grown in DMEM supplemented with 0.2 mg ml^{−1} geneticin for five days and the VLP-containing supernatant was collected, clarified at 500g for 15 min, concentrated to 50 ml, and supplemented with 10 µg ml^{−1} avidin and 0.1 M Tris–HCl (pH 8.0) before being passed through a 0.2-µm pore filter. The VLPs were then purified by streptactin-affinity chromatography and the eluate was concentrated and stored at −80°C.

For visualization by electron microscopy, MPRLV VLPs (80–100 ng µl^{−1}) were spotted on glow-discharged carbon grids, negatively stained with 2% uranyl acetate, analysed with a Tecnai G2 Bio-Twin electron microscope (FEI) and imaged with an Eagle camera (FEI).

Measurement of Gn/Gc–PCDH1 interaction by biolayer interferometry. The OctetRed 384 system (ForteBio, Pall LLC) was used to determine kinetic binding properties. Aminopropylsilane (APS) sensors were used to load VLPs (20 µg ml^{−1}) in PBS (pH 7.4) for 900 s until saturation was reached at 6 nM, washed for 60 s with PBS and quenched for 600 s in PBS–bovine serum albumin (BSA) (0.2 mg ml^{−1}). Subsequently, association of sEC1–2 was monitored for 600 s in PBS–BSA over the indicated concentration range, and dissociation was monitored for 300 s in PBS–BSA. Independent experiments were performed with different sensors to account for potential experimental artefacts resulting from intersensor variability. Non-specific interactions were monitored by applying the identical experimental set-up to empty APS sensors (PBS, pH 7.4 during the loading step). Global data fitting to a 1:1 binding model was used to estimate values for the k_{on} (association-rate constant), k_{off} (dissociation-rate constant) and K_d (equilibrium dissociation constant). The steady-state equilibrium concentration curve was fitted using a one-site-specific binding fit in Graphpad Prism.

Assay for sEC1–2-mediated inhibition of virus–cell attachment. HUVECs were seeded on six-well plates and chilled on ice. rVSVs bearing ANDV or HTNV Gn/Gc (1 µg total protein) were labelled with a lipophilic dye, FSL–fluorescein (Sigma-Aldrich; 5 µg ml^{−1}) as described¹⁴. Labelled viruses were preincubated with sEC1–2 (0–30 µg ml^{−1}; 0–1.1 µM) for 1 h at 37°C, and then allowed to attach to cells at a MOI of 1.5 IU per cell by centrifugation (2,500 r.p.m. for 60 min at 4°C) in HBSS (Corning). Cells were then placed on ice, washed extensively with cold HBSS to remove unbound virus, and fixed with 4% paraformaldehyde. Cell-surface-bound virus was analysed by flow cytometry as described¹⁴.

Assay for mAb-3305-mediated inhibition of virus–cell attachment. HUVECs, seeded on 24-well plates the previous day, were chilled on ice and incubated with mAb 3305 diluted in HUVEC medium at concentrations of 0–10 µg ml^{−1} (in serial three-fold dilutions) for 30 min on ice. rVSVs bearing ANDV or HTNV Gn/Gc (0.375 µg) were labelled with a lipophilic dye, 1,1'-diiododecyl-3,3',3'-tetramethylindodicarbocyanine 4-chlorobenzenesulfonate (DiD; 0.2–0.4 µM), by mixing and incubating the dye with the virus at 37°C for 1 h. Labelled viruses were added to the cells at an MOI of 0.4 IU per cell and allowed to attach by centrifugation (1,500g for 20 min at 4°C). Cells were placed on ice and washed extensively with cold HBSS (Corning) to remove unbound virus, collected from the plate, then fixed with 4% paraformaldehyde. Coverslips were mounted in Prolong with DAPI and imaged the same day by immunofluorescence microscopy. Virus and cell nuclei were detected and enumerated using Volocity (Perkin-Elmer) software's 'Find Objects' module. A total of 330–500 cells was analysed.

Animal welfare statement. Breeding, CRISPR–Cas9 genome engineering and challenge studies with Syrian golden hamsters were conducted under Institutional

Animal Care and Use Committee (IACUC)-approved protocols in compliance with the Animal Welfare Act, Public Health Service Policy, and other applicable federal statutes and regulations related to animals and experiments involving animals. The facilities where this research was conducted (Utah State University and US Army Medical Research Institute of Infectious Diseases (USAMRIID)) are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International (AAALAC), and adhere to principles stated in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011.

Generation of a *PCDH1*-KO Syrian hamster model. A portion of exon 2 of *PCDH1* was PCR-amplified from Syrian golden hamsters housed at Utah State University with primers MA-*PCDH1*-6723F (5'-TGCCTGTCGTTTACCCACC-3') and MA-*PCDH1*-7908R (5'-GGGAAAAGGAGCTTCCCAC-3') and subjected to Sanger sequencing. A panel of candidate sgRNAs was designed and assembled by overlapping PCR to generate human U6-promoter-driven sgRNA expression cassettes and screened for genome-editing efficiency in BHK21 baby hamster kidney cells stably expressing spCas9. The best candidate sgRNA (target sequence 5'-GGTAGTATACAAGGTGCCAGAGG-3'; PAM sequence is underlined), targeting exon 2 of the hamster *PCDH1* gene (nucleotides 7,063–7,085; GenBank accession number NW_004801621), was used for in vivo gene editing. This gRNA sequence was in vitro transcribed using the MEGAscript T7 Transcription kit (Thermo Fisher). Cas9 (PNA Bio; 2 µg) was incubated with sgRNA (1 µg) at room temperature for 15 min to generate sgRNA–Cas9 ribonucleoprotein (RNP) complexes, and then diluted to a concentration of 100 ng µl⁻¹ Cas9 and 50 ng µl⁻¹ sgRNA with 10 mM RNase-free TE buffer for pronuclear injections. Two- to three-month-old female Syrian hamsters (body weight 110–135 g) were super-ovulated by an intraperitoneal injection of pregnant mare's serum gonadotropin (PMSG; Sigma-Aldrich; 10–20 IU) at 09:00 on the day of post-oestrus discharge. The females were mated to fertile males at 19:00 on day 4 of the oestrous cycle and were humanely euthanized approximately 19 h later for zygote isolation. Zygotes were flushed from oviducts with warmed and equilibrated HECM-9 medium supplemented with 0.5 mg ml⁻¹ human serum albumin. Zygotes were then washed twice with HECM-9, transferred into 20-µl drops of HECM-9 covered by mineral oil in groups (of about 20) in a culture dish, and cultured at 37.5 °C under 10% CO₂, 5% O₂ and 85% N₂.

Cas9–sgRNA RNPs were injected into the zygote pronuclei in a dark room, and red filters were used on the microscope light source. A group of 15–20 hamster zygotes was transferred to a 100-µl HECM-9 drop in the microinjection dish, and 1–2 pl of Cas9–sgRNA RNP complex was injected into the male pronucleus of each zygote. After injection, zygotes were washed twice with equilibrated HECM-9 medium and incubated in HECM-9 medium covered by mineral oil for at least 30 min before embryo transfer. Embryos with normal morphology were bilaterally transferred into the oviducts of pseudopregnant recipients (10–15 embryos per oviduct) that were prepared by mating with vasectomized males at the same time that zygote donors were mated.

Genomic DNA was isolated from the pups at the age of 2 weeks, and a 595-base-pair (bp) product flanking the sgRNA target site was PCR-amplified by using primers MA-*PCDH1*-6723F (5'-TGCCTGTCGTTTACCCACC-3') and MA-*PCDH1*-7317R (5'-GCCATTCTGCACGAGTCTGT-3') and subjected to a T7 endonuclease I assay (NEB) to detect indels. Amplicons from pups bearing indels were topoisomerase (TOPO)-cloned and Sanger-sequenced to identify a founder female animal carrying a single-nucleotide (C) deletion at nucleotide 7,080 of the *PCDH1* gene. This deletion results in a frameshift at amino-acid position 58, leading to production of a truncated, 84-amino-acid polypeptide. Because it destroys a unique *BanI* restriction-enzyme site, a restriction-fragment-length polymorphism (RFLP) was used to genotype pups: the 595-bp wild-type allele can be digested by *BanI* into 343-bp and 252-bp products, whereas the edited allele is indigestible (see Extended Data Fig. 10a, b). Loss of *PCDH1* expression was confirmed by immunoblotting of lung homogenates from age-matched wild-type and *PCDH1*-knockout hamsters with mAb 3305 (Extended Data Fig. 10d).

Hamster challenge studies. Wild-type (Envigo) or *PCDH1*-knockout (Utah State University), 8–10-week-old male and female Syrian hamsters anaesthetized with isoflurane inhalation were exposed to a target dose of 2,000 PFU (actual dose determined by back-titration = 1,500 PFU) of ANDV strain Chile-9717869 diluted in PBS via the intranasal route, with 50 µl of the virus preparation delivered by pipette. Animals were observed daily for clinical signs of disease, morbidity and mortality. Moribund animals—described as being unresponsive or presenting with severe respiratory distress—were humanely euthanized on the basis of IACUC-approved criteria. Animal numbers were chosen so that the survival studies were adequately powered. Study personnel checking the animals were not blinded to treatment group; however, personnel were not privy to the details of specific treatments.

Histopathology and immunohistochemistry. Age-matched wild-type and *PCDH1*-knockout hamsters (one animal each) were humanely euthanized, and various

tissues—including lungs, heart, liver, spleen, kidneys, urinary bladder, brain, spinal cord, intestinal tract, testis and epididymis—were collected and preserved in 10% neutral-buffered formalin for 48 h at room temperature. Tissue were paraffin-embedded for sectioning (5–8 µm) and processed for haematoxylin–eosin staining for histopathological examination at the Utah Veterinary Diagnostic Laboratory. The presence of any pathological lesions was scored on a scale of 0–4, 0 indicating no lesions and 4 indicating severe lesions. No substantial microscopic lesions attributable to *PCDH1* loss were detected by the veterinary pathologist.

For histopathology of ANDV-challenged wild-type and *PCDH1*-knockout hamsters, lungs fixed in buffered formalin for 30 days, received from USAMRIID, were paraffin-embedded and 5-µm-thin sections were cut and stained with haematoxylin and eosin using standard procedures at the Histopathology and Comparative Pathology facility at Albert Einstein College of Medicine. Blinded slides were then scored on a scale of 1–4 for histopathological lesions by a veterinary pathologist using a previously described scoring metric²⁶.

Lung sections were also immunostained for viral antigen (ANDV nucleoprotein) as described³³. Briefly, deparaffinized and rehydrated tissue sections were subjected to antigen retrieval with 10 mM sodium citrate for 20 min in a steamer. Sections were then treated with 0.3% hydrogen peroxide for 10 min to block endogenous peroxidase activity and stained with a 1/1,000 dilution of an ANDV nucleoprotein-specific monoclonal antibody (clone 1A8F6, Austral Biologicals) for 1 h at room temperature, followed by ImmPress HRP anti-mouse IgG (peroxidase) polymer detection kit (Vector Labs) for 30 min at room temperature. After developing an immunohistochemical signal with the DAB peroxidase (HRP) substrate kit (Vector Labs), sections were counterstained with haematoxylin. Blinded slides were imaged by a veterinary pathologist.

Quantitative RT–PCR of hantavirus genomic RNA. RNA was purified from 0.25 ml of 10% lung homogenates or serum following inactivation with 0.75 ml of TRIzol LS reagent (Invitrogen), in accordance with the manufacturer's instructions. RNA concentrations were determined in a NanoDrop spectrophotometer (Thermo Fisher) and normalized to 40 ng µl⁻¹ in nuclease-free water. ANDV-specific quantitative reverse transcriptase (RT)–PCR reactions were completed with 250 ng of total RNA per 20-µl reaction, as described⁴¹. A standard curve of amplicon fluorescence intensity versus ANDV infectious titre (PFU per ml) was generated with serial 100-fold dilutions of RNA purified (as described above) from ANDV stock virus with a known infectious titre, and results were reported as genome equivalents per µg of total RNA.

Statistics and reproducibility. The *n* number associated with each dataset in the figures indicates the number of biologically independent samples. The number of independent experiments and the measures of central tendency and dispersion used in each case are indicated in the figure legends. The testing level (alpha) was 0.05 for all statistical tests. All analyses were carried out in GraphPad Prism. No statistical methods were used to predetermine sample size, except in Fig. 4a which used a power calculation (80% power, 5% type I error) to see a fourfold effect in survival. The experiments were not randomized.

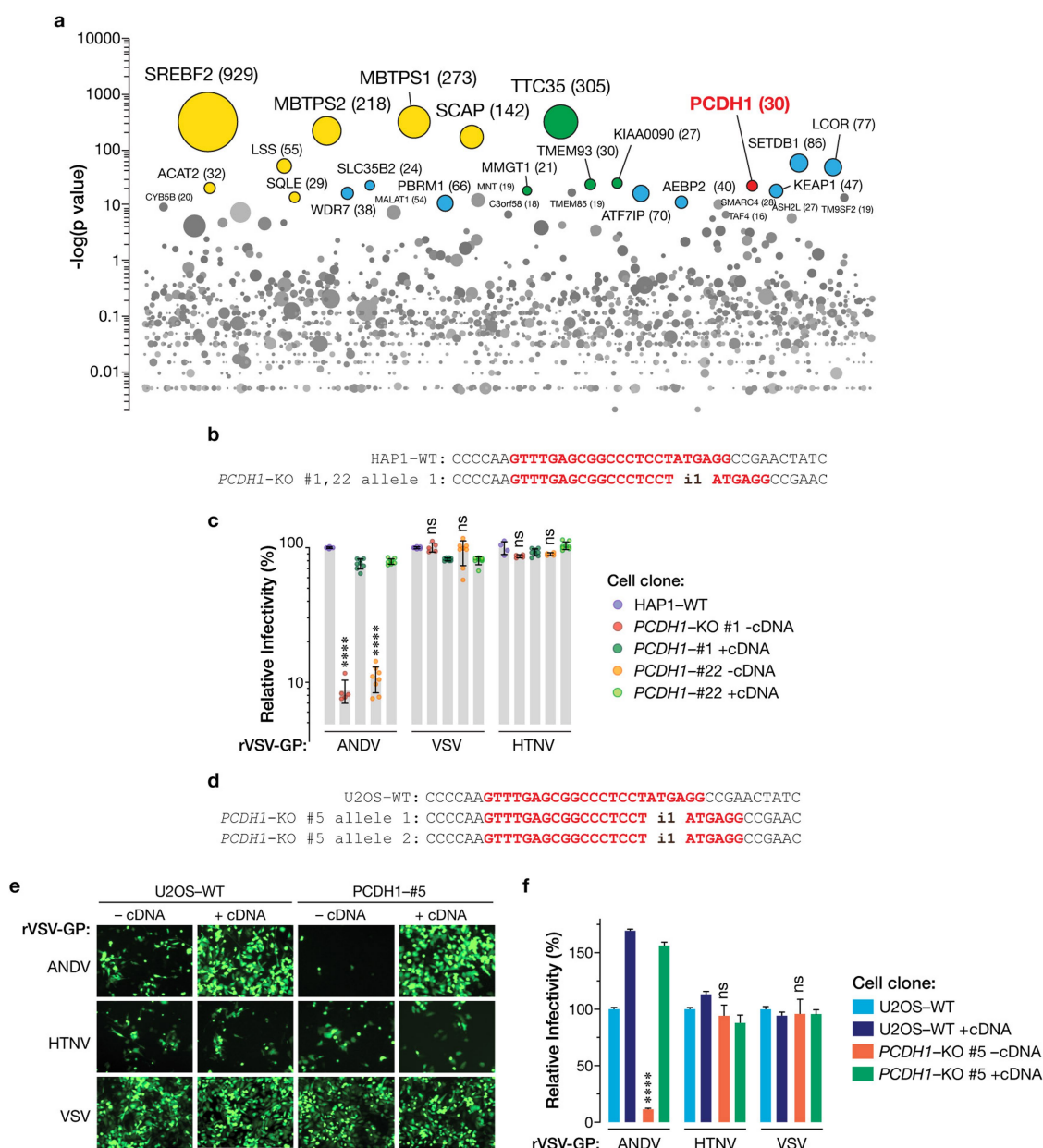
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Source Data for Figs. 1–4 are provided with the paper.

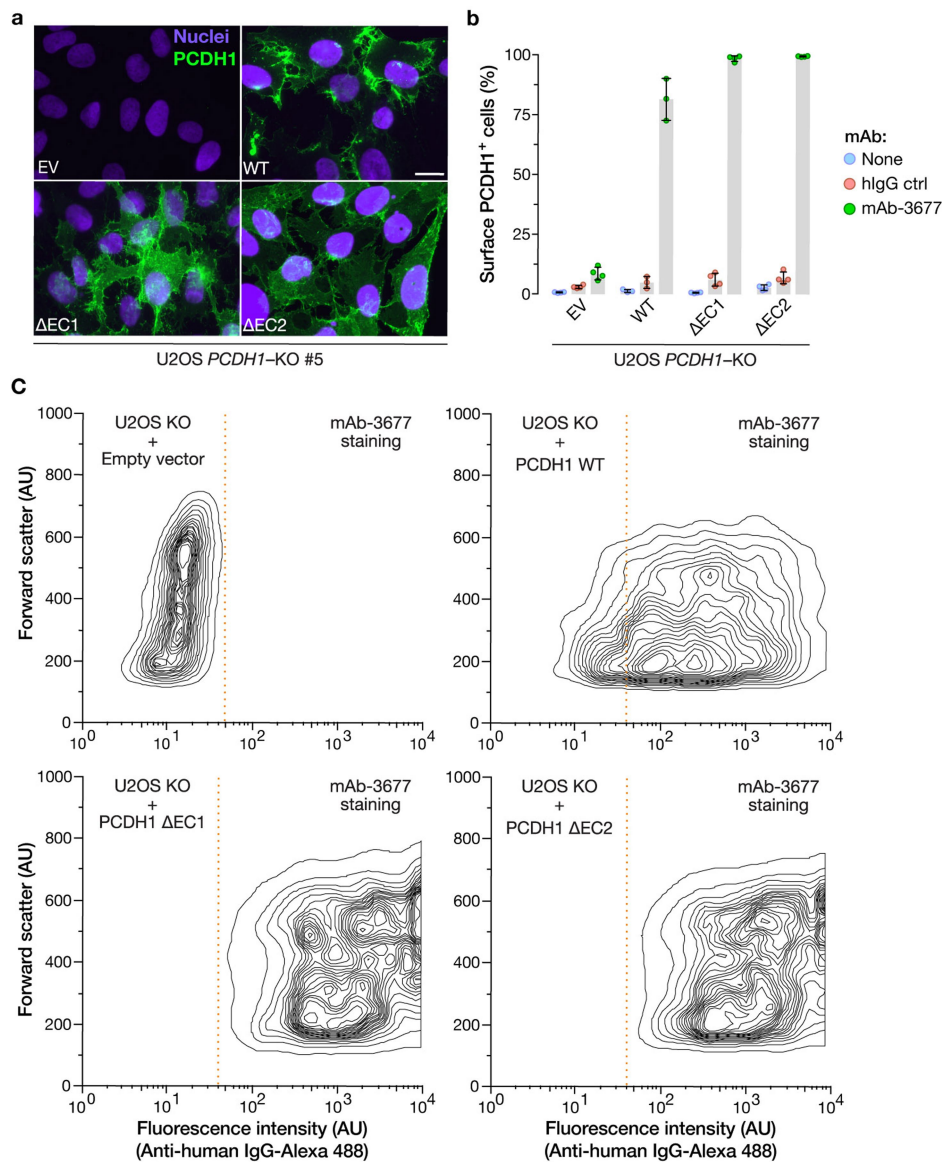
- Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
- Wong, A. C., Sandesara, R. G., Mulherkar, N., Whelan, S. P. & Chandran, K. A forward genetic strategy reveals destabilizing mutations in the Ebolavirus glycoprotein that alter its protease dependence during cell entry. *J. Virol.* **84**, 163–175 (2010).
- Kamentsky, L. et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).
- Hooper, J. W., Larsen, T., Custer, D. M. & Schmaljohn, C. S. A lethal disease model for hantavirus pulmonary syndrome. *Virology* **289**, 6–14 (2001).
- Lee, H. W., Lee, P. W. & Johnson, K. M. Isolation of the etiologic agent of Korean Hemorrhagic fever. *J. Infect. Dis.* **137**, 298–308 (1978).
- Schmaljohn, A. L. et al. Isolation and initial characterization of a newfound hantavirus from California. *Virology* **206**, 963–972 (1995).
- Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596 (1990).

35. Wec, A. Z. et al. A “Trojan horse” bispecific-antibody strategy for broad protection against ebolaviruses. *Science* **354**, 350–354 (2016).
36. Persson, H. et al. CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J. Mol. Biol.* **425**, 803–811 (2013).
37. Hornsby, M. et al. A high through-put platform for recombinant antibodies to folded proteins. *Mol. Cell. Proteomics* **14**, 2833–2847 (2015).
38. Cifuentes-Muñoz, N., Darlix, J. L. & Tischler, N. D. Development of a lentiviral vector system to study the role of the Andes virus glycoproteins. *Virus Res.* **153**, 29–35 (2010).
39. Hooper, J. W., Custer, D. M., Thompson, E. & Schmaljohn, C. S. DNA vaccination with the Hantaan virus M gene protects hamsters against three of four HFRS hantaviruses and elicits a high-titer neutralizing antibody response in Rhesus monkeys. *J. Virol.* **75**, 8469–8477 (2001).
40. Lefrançois, L. & Lyles, D. S. The interaction of antibody with the major surface glycoprotein of vesicular stomatitis virus. I. Analysis of neutralizing epitopes with monoclonal antibodies. *Virology* **121**, 157–167 (1982).
41. Trombley, A. R. et al. Comprehensive panel of real-time TaqMan polymerase chain reaction assays for detection and absolute quantification of filoviruses, arenaviruses, and New World hantaviruses. *Am. J. Trop. Med. Hyg.* **82**, 954–960 (2010).



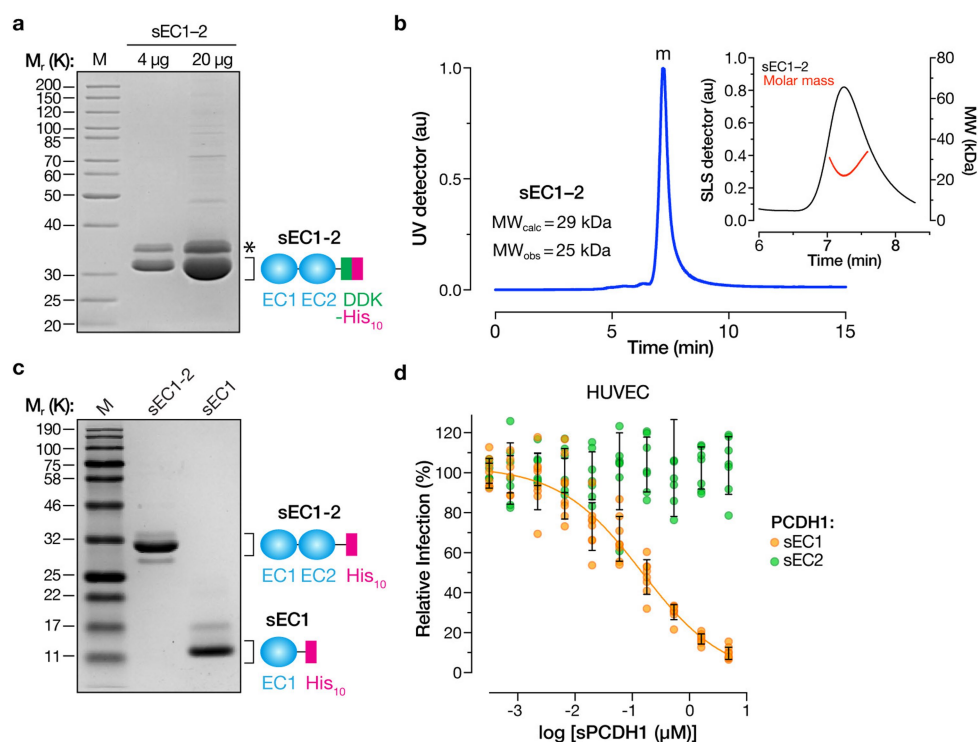
Extended Data Fig. 1 | *PCDH1* is required for entry of New World hantaviruses into HAP1 and U2OS cells. **a**, Genes enriched for gene-trap insertions in the rVSV-ANDV Gn/Gc-selected population versus unselected control cells. The size of each bubble reflects the number of independent gene-trap insertions observed. Candidate genes are associated with cholesterol metabolism (yellow), the endoplasmic reticulum membrane complex network (green), and transcription (blue). *PCDH1* (red) is a singleton hit. **b**, Single-cell HAP1 clones deficient for *PCDH1* were generated by CRISPR–Cas9 genome engineering. The sequences of *PCDH1*-knockout alleles in clones 1 and 22 are shown. The sgRNA target sequence is highlighted in red. i1, insertion of a single nucleotide in *PCDH1*. **c**, WT and HAP1 *PCDH1*-KO cell lines were exposed to rVSVs bearing the indicated viral glycoproteins at a MOI of 0.02 IU per cell. ‘+cDNA’ indicates complementation of *PCDH1*-KO cells with *PCDH1*. Cells were scored for infection at 20 hpi. One hundred per cent relative infection corresponds to 20–30% infected cells. Averages \pm s.d. are shown: HAP1-WT, three experiments, $n = 6$; *PCDH1*-

KO#1-cDNA, two experiments, $n = 5$; *PCDH1*-KO#1+cDNA, three experiments, $n = 8$; *PCDH1*-KO#22-cDNA, three experiments, $n = 8$, except for HTNV Gn/Gc, for which two experiments, $n = 4$; *PCDH1*-KO#1+cDNA, three experiments, $n = 8$; *PCDH1*-KO#22+cDNA, three experiments, $n = 8$. **d**, Single-cell U2OS clones deficient for *PCDH1* were generated by CRISPR–Cas9 genome engineering. The sequences of *PCDH1*-knockout alleles in clone 5 are shown. The sgRNA target sequence is highlighted in red. i1, insertion of a single nucleotide in *PCDH1*. **e**, **f**, WT and U2OS *PCDH1*-KO cell lines were exposed to the indicated rVSVs at a MOI of 0.02 IU per cell. Cells were scored for infection at 20 hpi. eGFP-positive infected cells (pseudocoloured green) were detected by fluorescence microscopy (**e**) and enumerated by automated counting (**f**). One hundred per cent relative infection corresponds to 10–20% infected cells. Averages \pm s.e.m. are shown, three experiments, $n = 13$. In **c**, **f**, wild type versus *PCDH1*-KO by two-way ANOVA with Tukey’s (**c**) or Sidak’s (**f**) tests: NS, $P > 0.05$; ****, $P < 0.0001$.



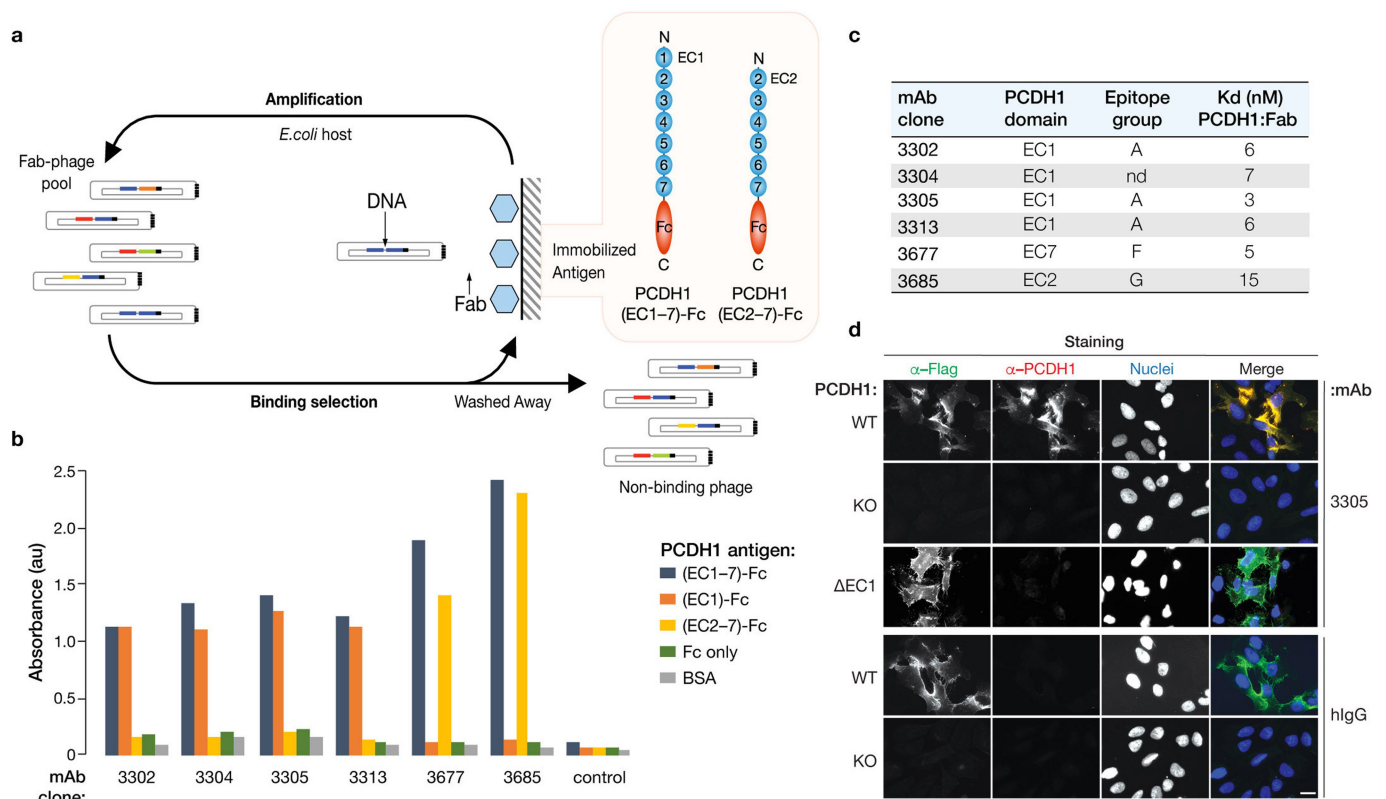
Extended Data Fig. 2 | Expression and plasma-membrane localization of *PCDH1* variants lacking domains EC1 or EC2 of *PCDH1*. **a**, U2OS *PCDH1*-KO cell lines complemented with the indicated *PCDH1* proteins were immunostained with an anti-Flag antibody and visualized by fluorescence microscopy. EV, empty vector. Scale bar, 20 μ m. **b**, **c**, Live cells from **a** were stained with the *PCDH1*-EC7-specific mAb 3677 at 4 $^{\circ}$ C to

detect cell-surface *PCDH1* and visualized by flow cytometry. Cells were gated on *PCDH1* immunofluorescence intensity (dotted red lines in **c**) to determine the percentage of cells with surface expression of each *PCDH1* protein. Averages \pm s.d. are shown in **b**: two experiments, $n = 4$, except in the case of WT, for which $n = 3$. **c**, Representative flow plots from **b**. Experiments were performed three times with similar results.



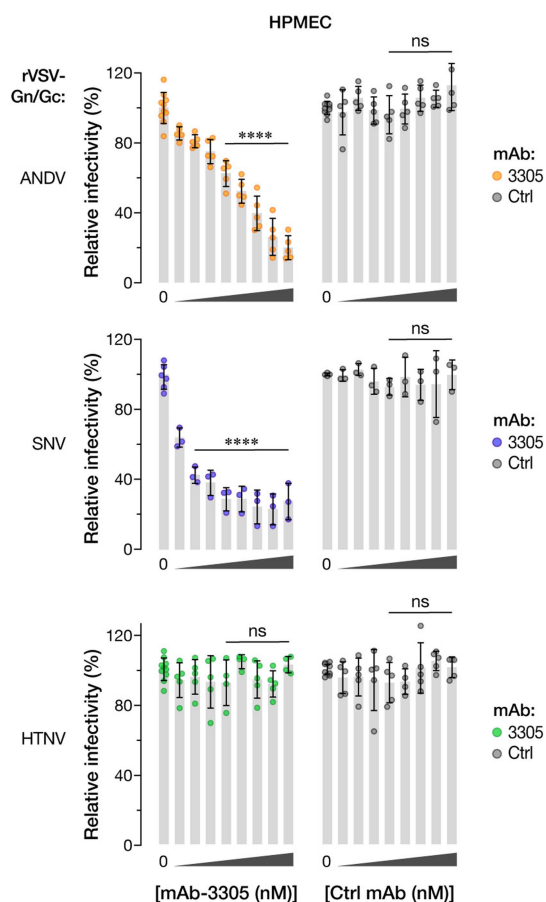
Extended Data Fig. 3 | Expression, purification and characterization of soluble PCDH1 proteins. **a**, Purified sEC1-2 bearing C-terminal Flag and decahistidine (His₁₀) epitope tags was resolved on an SDS-polyacrylamide gel and visualized by Coomassie blue staining. Asterisk, minor component of unknown origin. M_r , relative molecular weight (K denotes $\times 1,000$); M, monomer peak. The experiment was performed three times with similar results. **b**, SEC-MALS analysis of purified sEC1-2 from **a**. Absorbance (arbitrary units, au) was monitored at 280 nm; m, monomer peak. Calculated (MW_{calc}) and observed (MW_{obs}) molecular-weight estimates from MALS are shown in the inset. The experiment was performed twice

with similar results. **c**, Generation of purified sEC1, comprising the first extracellular cadherin domain of PCDH1, and bearing a C-terminal His₁₀ epitope tag. sEC1-2 is shown for comparison. The experiment was performed three times with similar results. **d**, Capacity of sEC1 and sEC2 to block hantavirus glycoprotein-dependent entry. rVSVs bearing ANDV Gn/Gc were preincubated with sEC1 or sEC2 (0–5 μ M, in serial threefold dilutions), and then allowed to infect HUVECs at a MOI of 0.2 IU per cell. Cells were scored for infection at 9 hpi. One hundred per cent relative infection corresponds to 10–20% infected cells. Averages \pm s.d.: three experiments, $n = 8$ for sEC1; $n = 7$ for sEC2.

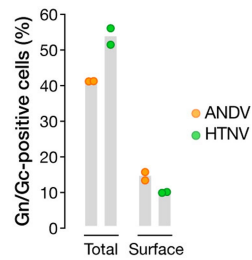


Extended Data Fig. 4 | Isolation of PCDH1-specific monoclonal antibodies. **a**, Flow chart showing isolation of PCDH1-specific mAbs from Library F by phage display. **b**, Capacity of selected mAb clones to recognize recombinant fusion proteins comprising PCDH1 ectodomains (containing or lacking EC1) and the Fc antibody domain. BSA, bovine serum albumin. Data from a representative ELISA experiment are shown. Experiments were performed three times with similar results. **c**, Kinetic binding analysis of selected Fabs to PCDH1(ectodomain)-Fc fusion proteins by surface plasmon resonance. nd, not determined; K_d , equilibrium dissociation

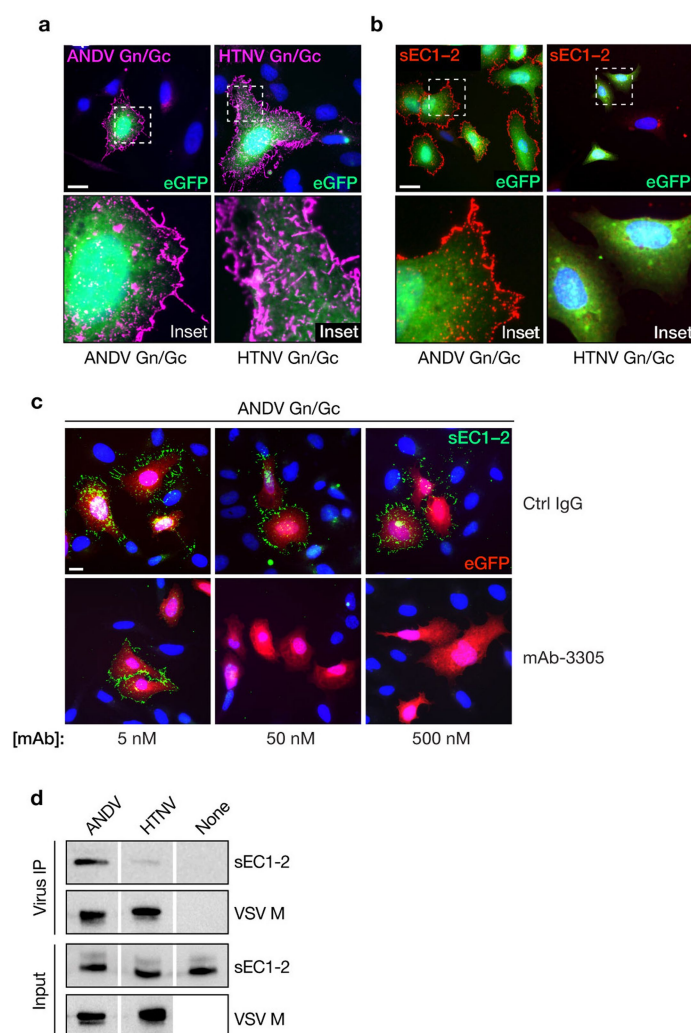
constant. **d**, mAb 3305 recognizes the first extracellular cadherin (EC1) domain of PCDH1. U2OS *PCDH1*-KO cells, uncomplemented (KO) or complemented either with full-length PCDH1 (WT) or with a PCDH1 variant lacking the EC1 domain (Δ EC1), were co-immunostained with anti-Flag (α -Flag) mAb and anti-PCDH1 mAb 3305, or with anti-Flag and negative control antibodies (hIgG). Cells were visualized by fluorescence microscopy. Scale bar, 20 μ m. Data from a representative experiment are shown. Experiments were performed three times with similar results.



Extended Data Fig. 5 | PCDH1 EC1-specific mAb 3305 blocks ANDV and SNV entry into primary HPMECs. HPMECs were preincubated with mAb 3305 or with human IgG control (Ctrl) ($0\text{--}100\text{ }\mu\text{g ml}^{-1}$, $0\text{--}680\text{ nM}$, in serial threefold dilutions), and then exposed to the indicated rVSVs Gn/Gc glycoproteins at a MOI of 0.2 IU per cell. Infected cells were scored at 9 hpi. One hundred per cent relative infectivity corresponds to 15–20% infected cells. Averages \pm s.d.: three experiments; no-mAb samples ('0'), $n = 10$ for ANDV and HTNV, $n = 3$ for SNV; mAb-treated samples, $n = 5$ for ANDV and HTNV, $n = 3$ for SNV. mAb 3305 versus control mAb, two-way ANOVA with Dunnett's test: ns, $P > 0.05$; **** $P < 0.0001$.

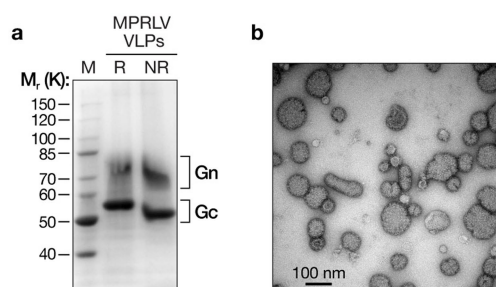


Extended Data Fig. 6 | The monoclonal antibody 1E11/D3 recognizes HTNV and ANDV Gn/Gc. 293FT cells transfected with plasmids encoding ANDV or HTNV Gn/Gc were analysed by flow cytometry using 1E11/D3 mAb for total (after permeabilization) and surface (without permeabilization) expression of Gn/Gc. Empty-vector-transfected cells were used as a negative control for gating. Results are from a representative experiment with $n = 2$. Experiments were performed five times for ANDV and twice for HTNV with similar results.



Extended Data Fig. 7 | PCDH1 mediates viral entry by direct binding to hantavirus glycoproteins. Capacity of hantavirus glycoproteins expressed at the cell surface to capture Flag-tagged sEC1-2 from solution. **a**, rVSVs bearing ANDV or HTNV Gn/Gc were allowed to infect U2OS cells, and cell-surface expression of Gn/Gc was detected by immunofluorescence microscopy using mAb 1E11/D3. **b**, Cells expressing Gn/Gc were then exposed to sEC1-2 (200 nM), and sEC1-2 binding to the cell surface was detected by immunofluorescence microscopy using an anti-Flag mAb. **c**, The capacity of PCDH1-specific mAb 3305 to block binding of sEC1-2 to Gn/Gc-expressing cells was determined as in **b**. sEC1-2 (50 nM) was preincubated with the indicated amounts of a control IgG or mAb 3305,

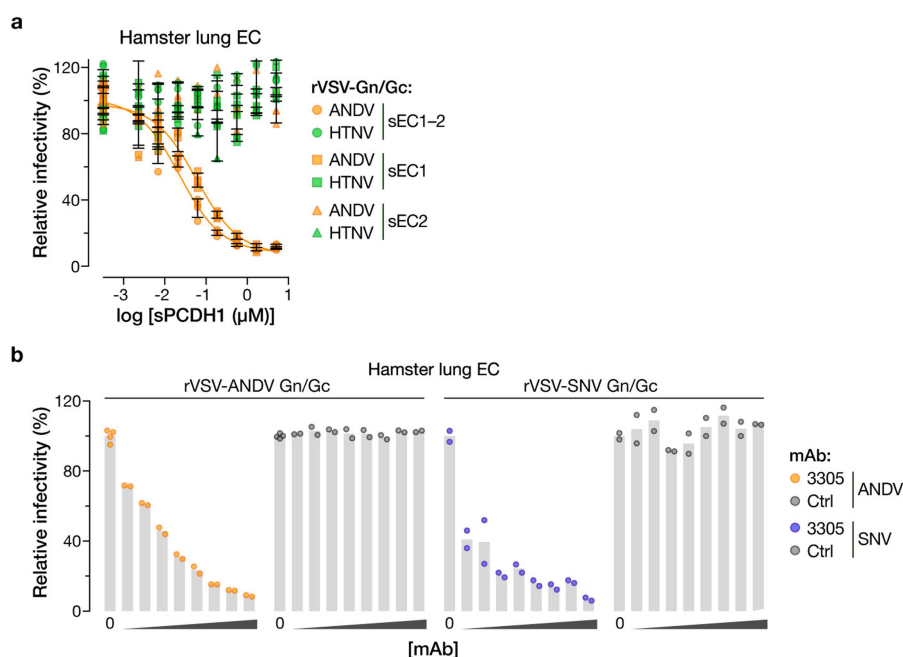
and then exposed to Gn/Gc-expressing cells. In **a–c**, representative images are shown, and experiments were performed three times with similar results. Scale bars, 20 μ m. **d**, Biotinylated rVSVs bearing ANDV or HTNV Gn/Gc were incubated with sEC1-2 and then captured with streptavidin magnetic beads. Co-precipitated viral particles and sEC1-2 ('Virus IP') and a fraction of the input material ('Input') were detected by immunoblotting with mAb 23H12, specific for the VSV M matrix protein, and an anti-Flag mAb, respectively. 'None' indicates control precipitation of sEC1-2 in the absence of viral particles. Representative images are shown. Experiments were performed three times with similar results.



c

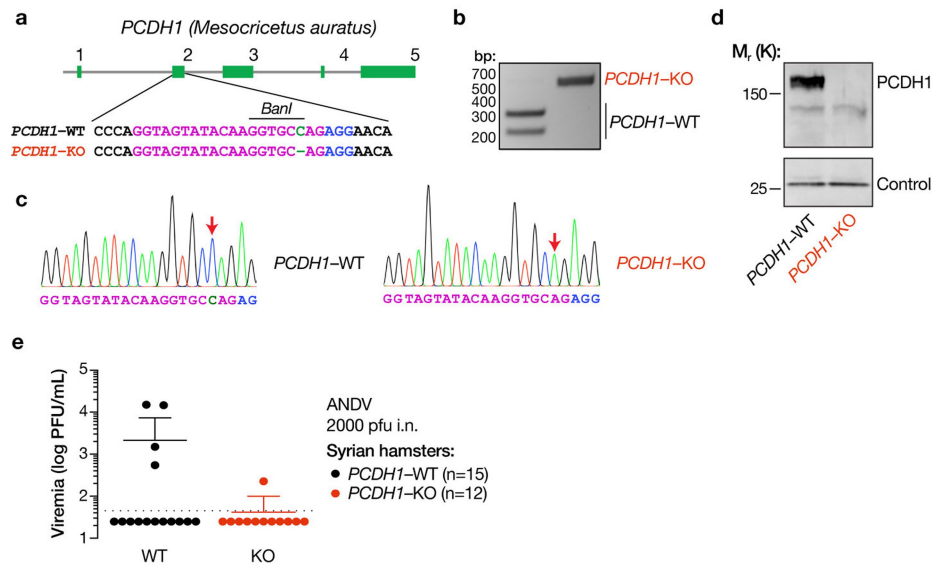
MPRLV VLPs	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_D (M)
PCDH1 sEC1-2	$(2.76 \pm 0.01) \times 10^5$	$(1.73 \pm 0.06) \times 10^{-3}$	$(6.26 \pm 0.03) \times 10^{-9}$
PCDH1 sEC1-2	$(2.00 \pm 0.01) \times 10^5$	$(1.38 \pm 0.05) \times 10^{-3}$	$(6.91 \pm 0.03) \times 10^{-9}$

Extended Data Fig. 8 | Expression and purification of MPRLV VLPs and their interaction with soluble PCDH1. **a**, Purified MPRLV VLPs (8 μ g protein) bearing an N-terminal StrepTag were resolved on an SDS-polyacrylamide gel and visualized by Coomassie blue staining. R, reducing conditions; NR, nonreducing conditions. **b**, MPRLV VLPs were negatively stained with uranyl acetate and visualized by electron microscopy. **c**, Kinetic constants for MPRLV VLP-sEC1-2 interaction were determined by biolayer interferometry. k_{on} , association-rate constant; k_{off} , dissociation rate constant. Values \pm 95% confidence intervals derived from curve fits are shown for two independent experiments. These experiments were performed three times with similar results.



Extended Data Fig. 9 | PCDH1-EC1 is required for hantavirus Gn/Gc-dependent entry and infection in primary Syrian hamster lung endothelial cells. **a**, Capacity of soluble PCDH1 (sPCDH1) variants (sEC1-2, sEC1 or sEC2) to block hantavirus glycoprotein-dependent entry. rVSVs bearing ANDV or HTNV Gn/Gc glycoproteins were preincubated with the indicated amounts of the indicated sPCDH1 variants (0–5 μM , in serial threefold dilutions) at room temperature for 1 h, and then allowed to infect Syrian hamster lung endothelial cells (ECs). One hundred per cent relative infectivity corresponds to 15–20% infected

cells. Averages \pm s.d.: two experiments, $n = 4$. **b**, Capacity of PCDH1-EC1-specific mAb 3305 to block hantavirus glycoprotein-dependent entry. Syrian hamster lung ECs were preincubated with mAb 3305 or control human IgG (0–100 $\mu\text{g ml}^{-1}$, 0–680 nM, in serial threefold dilutions), and then exposed to rVSVs bearing ANDV or SNV Gn/Gc at an MOI of 0.2 IU per cell. Cells were scored for infection at 9 hpi. One hundred per cent relative infectivity corresponds to 15–20% infected cells. Two experiments, $n = 2$, except in the case of no mAb ('0') for which $n = 4$.



Extended Data Fig. 10 | Generation of a *PCDH1*-KO knockout Syrian hamster by CRISPR-Cas9 genome engineering. **a**, Organization of the *PCDH1* gene of the Syrian hamster (*M. auratus*). The sequence in exon 2 of *PCDH1* that was targeted by an sgRNA is shown in magenta. Knockout animals bear two *PCDH1* alleles that have been edited to lack a single nucleotide, highlighted in green. The sgRNA PAM is highlighted in blue. **b**, A PCR-RFLP strategy, based on loss of digestion by the restriction endonuclease *BanI*, was used to detect genome editing and genotype animals. **c**, PCR-RFLP results were confirmed by Sanger DNA sequencing of PCR amplicons from WT and genome-edited animals. Sequencing

traces are shown. Sequence features are highlighted as in **a**. Red arrows denote the site of gene editing in the *PCDH1*-KO allele. Experiments were performed twice with similar results. **d**, Lung tissue isolated from WT and *PCDH1*-KO hamsters was solubilized, normalized by protein content, and subjected to SDS-PAGE. PCDH1 was detected by immunoblotting with EC1-specific mAb 3305. Control, nonspecific loading control. Experiments were performed three times with similar results. **e**, Viral loads in the sera of WT and *PCDH1*-KO hamsters at 14 dpi. The limit of detection is shown as a dotted line. Experiments in **b–e** were performed twice with similar results.

The metabolite BH4 controls T cell proliferation in autoimmunity and cancer

Shane J. F. Cronin^{1,2,3}, Corey Seehus^{2,3}, Adelheid Weidinger⁴, Sebastien Talbot^{2,3,5}, Sonja Reissig⁶, Markus Seifert⁷, Yann Pierson⁸, Eileen McNeill^{9,10}, Maria Serena Longhi¹¹, Bruna Lenfers Turnes¹², Taras Kreslavsky^{13,14}, Melanie Kogler¹, David Hoffmann¹, Melita Ticevic¹, Débora da Luz Scheffer¹², Luigi Tortola¹, Domagoj Cikes¹, Alexander Jais¹⁵, Manu Rangachari^{16,17}, Shuan Rao¹, Magdalena Paolino¹⁴, Maria Novatchkova¹³, Martin Aichinger¹⁴, Lee Barrett^{2,3}, Alban Latremoliere¹⁸, Gerald Wirnsberger¹⁹, Guenther Lametschwandtner¹⁹, Meinrad Busslinger¹³, Stephen Zicha²⁰, Alexandra Latini^{2,3,12}, Simon C. Robson^{9,10}, Ari Waisman⁶, Nick Andrews^{2,3}, Michael Costigan^{2,3,21,22}, Keith M. Channon^{9,10}, Guenter Weiss⁷, Andrey V. Kozlov⁴, Mark Tebbe¹⁷, Kai Johnsson^{8,23}, Clifford J. Woolf^{2,3*} & Josef M. Penninger^{1*}

Genetic regulators and environmental stimuli modulate T cell activation in autoimmunity and cancer. The enzyme co-factor tetrahydrobiopterin (BH4) is involved in the production of monoamine neurotransmitters, the generation of nitric oxide, and pain^{1,2}. Here we uncover a link between these processes, identifying a fundamental role for BH4 in T cell biology. We find that genetic inactivation of GTP cyclohydrolase 1 (GCH1, the rate-limiting enzyme in the synthesis of BH4) and inhibition of sepiapterin reductase (the terminal enzyme in the synthetic pathway for BH4) severely impair the proliferation of mature mouse and human T cells. BH4 production in activated T cells is linked to alterations in iron metabolism and mitochondrial bioenergetics. In vivo blockade of BH4 synthesis abrogates T-cell-mediated autoimmunity and allergic inflammation, and enhancing BH4 levels through GCH1 overexpression augments responses by CD4- and CD8-expressing T cells, increasing their antitumour activity in vivo. Administration of BH4 to mice markedly reduces tumour growth and expands the population of intratumoral effector T cells. Kynurenine—a tryptophan metabolite that blocks antitumour immunity—inhibits T cell proliferation in a manner that can be rescued by BH4. Finally, we report the development of a potent SPR antagonist for possible clinical use. Our data uncover GCH1, SPR and their downstream metabolite BH4 as critical regulators of T cell biology that can be readily manipulated to either block autoimmunity or enhance anticancer immunity.

GCH1—the first enzyme in the de novo BH4-synthesis pathway—is known to be expressed in activated T cells^{3,4}. Using isolated CD4⁺ and CD8⁺ T cells from a *Gch1-Gfp* reporter mouse line¹ (in which *Gfp* encodes green fluorescent protein), we confirmed that GCH1 is induced in activated T cells in response both to phorbol myristate acetate (PMA) and ionomycin, and to stimulation of T cell receptors (TCRs) by anti-CD3/CD28 antibodies (Extended Data Fig. 1a–c). To explore the function of the GCH1–BH4 pathway in these cells, we generated mice in which *Gch1* is knocked out specifically in T cells by crossing *Lck-cre* driver mice with *Gch1^{fl/fl}* (ref. ⁵) mice (producing *Gch1;Lck* animals). These *Gch1;Lck* mice showed normal numbers

of thymic and peripheral T cells compared with Cre-only controls (Extended Data Fig. 1d); that is, lack of GCH1 does not influence T cell development or peripheral T cell homeostasis. Stimulation of mature peripheral CD4⁺ T cells from *Gch1;Lck* mice revealed, as expected, severely reduced GCH1 protein and BH4 production relative to controls (Fig. 1a, b). Shortly after TCR engagement (at 16 hours), we observed no differences between *Gch1;Lck* and control T cells in either the expression of surface activation markers or the secretion of interleukin (IL)-2 (Fig. 1c, d). Similar results were obtained with CD8⁺ T cells (data not shown). However, TCR-stimulated *Gch1*-deficient CD4⁺ and CD8⁺ T cells did display markedly reduced proliferation (Fig. 1e, f and Extended Data Fig. 1e, f). In contrast to this antiproliferative effect on peripheral T cells, *Gch1* ablation did not affect the proliferation of DN3a thymocytes co-cultured with OP9–DL1 stromal cells (Extended Data Fig. 1g–i). Moreover, there were no obvious differences in the survival of thymocytes or of mature naive peripheral T cells (Extended Data Fig. 2a, b).

To validate these findings, we crossed a different T-cell-specific Cre mouse line, *RORgammat cre*⁶ (which expresses Cre recombinase under the *Rorc* promoter), with *Gch1^{fl/fl}* (ref. ⁵) mice. Again, specific loss of GCH1 in T cells did not affect thymocyte development or peripheral T cell homeostasis (data not shown). However, we again found that GCH1 is a key regulator of mature T cell proliferation (Fig. 1e and Extended Data Fig. 2c, d). B-cell-specific deletion of *Gch1* using *MB1-cre*⁷ (*MB1-cre* is also known as *Cd79a^{tm1(cre)Reth}*) did not affect B cell development or function (Extended Data Fig. 2e–h). Moreover, loss of GCH1 had no effect on the development, numbers or suppressive capacity of peripheral regulatory T cells (Extended Data Fig. 3a–f). We conclude that GCH1 induction and BH4 synthesis are required for effective proliferation of CD4⁺ and CD8⁺ T cells.

To investigate whether *Gch1*-ablated, BH4-deficient T cells are defective in vivo, we studied several models of T-cell-dependent inflammation. In a colitis model in which naive, CD4⁺ T cells are transferred into hosts that lack recombination-activating gene 1 (*Rag1*)⁸, transfer of *Gch1*-deficient CD4⁺ T cells resulted in a substantially lower influx of immune cells into the gut, with less colonic inflammation and colitis

¹IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna, Austria. ²Department of Neurobiology, Harvard Medical School, Boston, MA, USA. ³FM Kirby Neurobiology Center, Boston Children's Hospital, Boston, MA, USA. ⁴Ludwig Boltzmann Institute for Experimental and Clinical Traumatology, AUVA Research Center, Vienna, Austria. ⁵Département de Pharmacologie et Physiologie, Université de Montréal, Montréal, Québec, Canada. ⁶Institute for Molecular Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany. ⁷Department of Internal Medicine II (Infectious Diseases, Immunology, Rheumatology and Pneumology), Medical University of Innsbruck, Innsbruck, Austria. ⁸Institute of Chemical Sciences and Engineering, Institute of Bioengineering, National Centre of Competence in Research (NCCR) in Chemical Biology, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. ⁹Division of Cardiovascular Medicine, British Heart Foundation Centre for Research Excellence, John Radcliffe Hospital, University of Oxford, Oxford, UK. ¹⁰Wellcome Trust Centre for Human Genetics, Roosevelt Drive, University of Oxford, Oxford, UK. ¹¹Division of Gastroenterology and Liver Center, Department of Medicine, Beth Israel Deaconess Medical Center (BIDMC) and Harvard Medical School (HMS), Harvard University, Boston, MA, USA. ¹²LABOX, Departamento de Bioquímica, Universidade Federal de Santa Catarina, Florianópolis, Brazil. ¹³Research Institute of Molecular Pathology, Vienna Biocenter, Campus-Vienna-Biocenter 1, Vienna, Austria. ¹⁴Karolinska Institute, Department of Medicine Solna, Center for Molecular Medicine, Karolinska University Hospital Solna, Stockholm, Sweden. ¹⁵Department of Neuronal Control of Metabolism, Max Planck Institute for Metabolism Research, Cologne, Germany. ¹⁶Department of Neurosciences, Centre de Recherche de CHU de Québec-Université Laval, Québec, Québec, Canada. ¹⁷Department of Molecular Medicine, Faculty of Medicine, Université Laval, Québec, Québec, Canada. ¹⁸Neurosurgery Department, Johns Hopkins School of Medicine, Baltimore, MD, USA. ¹⁹Apeiron Biologics AG, Vienna, Austria. ²⁰Quartet Medicine, 400 Technology Square, Cambridge, MA, USA. ²¹Department of Anesthesia, Harvard Medical School, Boston, MA, USA. ²²Boston Children's Hospital, Boston, MA, USA. ²³Department of Chemical Biology, Max-Planck Institute for Medical Research, Heidelberg, Germany. *e-mail: clifford.woolf@childrens.harvard.edu; josef.penninger@imba.oew.ac.at

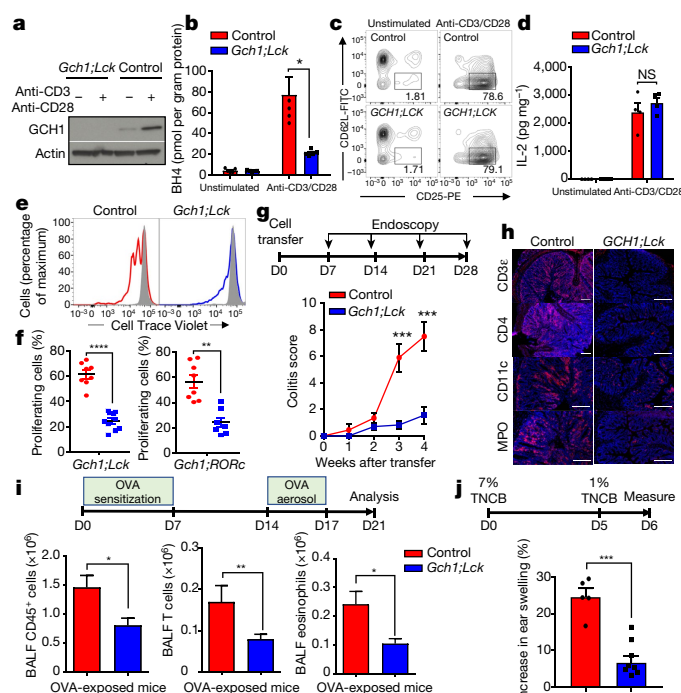


Fig. 1 | The BH4 pathway is indispensable for effective T cell proliferation in vitro and in vivo. **a**, Immunoblot of GCH1 after 24 hours of TCR stimulation with anti-CD3/CD28 antibodies in CD4⁺ T cells. Experiment repeated three times with similar results. Staining for actin acts as a control. **b**, BH4 production upon 24 hours of anti-CD3/CD28 stimulation in purified CD4⁺ control and *Gch1*-null T cells. Individual data (dots and squares; $n = 5$ mice in each case) are shown as means \pm s.e.m. **c**, **d**, Representative fluorescence-activated cell sorting (FACS) plot depicting early activation markers (CD62L and CD25; **c**) and IL-2 secretion (**d**) before and after T cell stimulation (16 h). $n = 5$ independent samples. Experiment repeated two independent times with similar results. FITC, fluorescein isothiocyanate. Naive T cells, CD25^{low}, CD62L^{high}; activated T cells, CD25^{high}, CD62L^{low}. **e**, Proliferation of CD4⁺ T cells after three days of stimulation of control and *Gch1*;Lck mice. Cell Trace Violet gets diluted in proliferating cells (see Methods). Representative data are shown from more than 15 experiments with similar results. **f**, Quantification of CD4⁺ T cell proliferation from individual *Gch1*;Lck (left; $n = 10$) and *Gch1*;RORc (right; $n = 7$) mice. **g**, **h**, Transfer colitis model of intestinal autoimmunity. **g**, Schematic outline (top) and colitis scores of transferred control and *Gch1*-ablated CD4⁺ T cells into *Rag1*^{-/-} hosts (bottom). D, day. $n = 10$ mice. **h**, Representative immunofluorescence depicting intestinal infiltration of various immune cells (CD3⁺, CD4⁺, CD11c⁺ and myeloperoxidase (MPO)⁺ cells). Scale bar, 200 μ m. **i**, Allergic airway inflammatory disease model (top) and quantification of inflammatory cells in bronchoalveolar lavage fluids (BALFs; bottom). $n = 35$ control mice; $n = 31$ *Gch1*;Lck mice. OVA, ovalbumin. **j**, Percentage increase in ear swelling after re-challenge using the 2,4,6-trinitrochlorobenzene (TNCB)-dependent skin hypersensitivity model. $n = 8$ control mice; $n = 9$ *Gch1*;Lck mice. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (two-tailed Student's *t*-test for **b**, **d**, **f**, **i**, **j**; two-way analysis of variance (ANOVA) with Dunnett's comparison for **g**). Data are mean \pm s.e.m.

development (Fig. 1g, h). Although numbers of colonic and mesenteric lymph node CD4⁺ T cells were greatly reduced, the production of the inflammatory T cell cytokines IL-17A and interferon (IFN)- γ were apparently not affected by selective *Gch1* deletion (Extended Data Fig. 3g–i). Next, we used a model of type-2 allergic airway inflammation in which immune cells—particularly CD4⁺ T helper-2 cells and eosinophils—are central to disease pathology^{9,10}. Compared with controls, *Gch1*;Lck mice showed substantially fewer CD45⁺ cells, eosinophils and T cells in bronchoalveolar lavage (Fig. 1i). Moreover, T-cell-dependent ovalbumin-induced immune responses were reduced during primary immunization and re-challenge (Extended Data Fig. 4a). *Gch1*;Lck mice also showed greatly reduced inflammatory

responses in a T-cell-mediated skin dermatitis model¹¹ (Fig. 1j) and in the experimental autoimmune encephalomyelitis (EAE) model of multiple sclerosis^{12,13} (Extended Data Fig. 4b, c). Therefore, genetic ablation of *Gch1* in T cells alleviates T-cell-mediated inflammatory intestinal, airway, skin and brain diseases.

Inhibiting GCH1 pharmacologically is challenging because of its inaccessible active sites^{14,15}. Therefore, we used an inhibitor, SPRI3, that targets sepiapterin reductase (SPR)—the terminal enzyme in the de novo BH4 synthesis pathway (Fig. 2a and Extended Data Fig. 4d). Purified naive CD4⁺ T cells treated with SPRI3 showed lower BH4 levels than did vehicle-treated cells following TCR stimulation (Fig. 2a). SPRI3-treated, TCR-stimulated CD4⁺ and CD8⁺ cells displayed a defect in proliferation that was similar to that of T cells in which *Gch1* was genetically ablated (Fig. 2b), without affecting the survival of non-stimulated T cells or the induction of early activation markers (Extended Data Fig. 4e, f). Pulse labelling with 5-ethynyl-2'-deoxyuridine (EdU) revealed that SPRI3 treatment and *Gch1* deficiency resulted in substantially fewer S-phase cells after TCR stimulation than did vehicle treatment, culminating in increased cell death (Extended Data Fig. 4g). In vivo, SPRI3 administration significantly ameliorated colitis, greatly diminishing the intestinal infiltration of T cells and other immune cells after CD4⁺ T cell transfer (Fig. 2c). SPRI3 treatment also reduced immune-cell infiltration into the lungs after challenge involving inhaled ovalbumin in sensitized mice (Fig. 2d). To determine whether these findings translate to human T cells, we isolated human peripheral blood mononuclear cells from different healthy donors ($n = 4$). Following anti-CD3/CD28 stimulation, SPRI3-treated, freshly isolated human T cells also exhibited greatly reduced proliferation compared with vehicle-treated cells (Fig. 2e). Moreover, we observed a substantial decrease in proliferative capacity in SPRI3-treated purified human effector CD4⁺ T cells after anti-CD3/CD8 re-stimulation (Fig. 2f).

To explore the molecular mechanisms responsible for the proliferation deficit, we carried out gene-expression profiling in TCR-stimulated CD4⁺ T cells from control and *Gch1*;Lck mice. Analysis of the greatly altered genes confirmed that loss of GCH1 did not affect early T cell activation (data accessible through Gene Expression Omnibus (GEO), accession number GSE108101 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108101>)). Biogenic amines or their amino-acid precursors—the production of which involves BH4 as a co-factor²—were also unaffected (Extended Data Fig. 4j, k and Supplementary Table 1). Several genes involved in iron homeostasis or its availability were upregulated in the absence of GCH1—a finding that was confirmed by western blot of activated *Gch1*-ablated T cells (Fig. 2g). Total iron levels were greatly reduced in TCR-activated *Gch1*-ablated CD4⁺ T cells compared with control cells (Fig. 2h). As one of the most upregulated genes in *Gch1*-ablated cells was mitoferrin—a mitochondrial iron transporter—and because iron is critical for mitochondrial respiration¹⁶, we analysed the energy needs of *Gch1*-deficient activated T cells. Both *Gch1*-deficient and SPRI3-treated CD4⁺ T cells synthesized less ATP than control cells after anti-CD3/CD28 stimulation (Extended Data Fig. 5a, b). Furthermore, levels of both lactate and pyruvate were enhanced in activated *Gch1*-deficient T cells, indicating augmented glycolysis (Extended Data Fig. 5c), and suggesting that the loss of GCH1 expression affects mitochondrial respiration.

Following anti-CD3/CD28 stimulation, mitochondrial respiration and oxygen consumption were much lower in BH4-deficient T cells than in control cells (Fig. 2i and Extended Data Fig. 5d–g). Cytochrome *c*—a redox-active protein containing haem groups that reversibly alternate between their Fe²⁺ and Fe³⁺ oxidation states—is important for the mitochondrial electron-transport chain (ETC), and we confirmed earlier reports^{17–19} that BH4 efficiently reduces ferri (Fe³⁺)-cytochrome *c* to ferro (Fe²⁺)-cytochrome *c* at doses that are physiological, in activated T cells (Fig. 2j). Critically, we could rescue ETC function by providing reduced cytochrome *c* to BH4-deficient cells (Fig. 2k, l). Moreover, impaired ETC in activated *Gch1*-ablated and SPRI3-treated CD4⁺ T cells was associated with elevation of superoxide reactive oxygen species (ROS; Extended Data Fig. 6a, b). The superoxide scavenger *N*-acetylcysteine (NAC) only partially rescued the proliferation defect

of *Gch1*-ablated T cells, and NAC addition did not rescue the iron deficiency observed in activated *Gch1*-ablated CD4⁺ T cells, nor did it enhance ATP production (Extended Data Fig. 6c–f), suggesting that the enhanced ROS are the result of mitochondrial dysregulation. BH4 is a co-factor for nitric oxide synthase (NOS) enzymes and is required for nitric oxide (NO) production². However, under our experimental conditions we did not observe detectable expression of inducible NOS or NO production until several days after T cell activation (Extended Data Fig. 6g–j). Our data indicate that antigen-receptor-stimulated, BH4-depleted T cells display a defective iron-redox cycling of cytochrome c, which leads to mitochondrial dysfunction.

Given that SPRi3 has relatively low potency and a short half-life, we developed a novel SPR inhibitor, QM385, which is structurally distinct from SPRi3 (Fig. 3a). QM385 binds with high affinity to human SPR in a cell-free assay, and efficiently reduced BH4 levels in anti-CD3/CD28 activated mouse splenocytes and in anti-CD3/CD28 activated human peripheral blood mononuclear cells (Extended Data Fig. 7a–c). QM385 is orally bioavailable, has a long half-life (Supplementary Table 2) and dose-dependently reduces plasma levels of BH4 while concurrently increasing levels of sepiapterin (Fig. 3b), a sensitive biomarker of SPR inhibition¹. QM385 did not inhibit a panel of physiologically important targets or closely related reductases (Supplementary Table 3), and in vivo administration did not result in detectable adverse effects. QM385 treatment resulted in markedly less proliferation of CD4⁺ T cells in vitro (Fig. 3c). Moreover, in activated CD4⁺ T cells, QM385 substantially reduced ETC function and ATP levels, and led to an elevation in superoxide ROS (Extended Data Fig. 7d–f). Altogether, QM385 phenocopies the effects of SPRi3 in vitro, albeit at much lower concentrations. Importantly, oral administration of QM385 to mice for three days greatly reduced the number of inflammatory T cells and eosinophils in the ovalbumin-induced and house dust mite (HDM) airway allergic inflammation models (Fig. 3d and Extended Data Fig. 7g), which are T cell dependent²⁰. QM385 effectively inhibited the proliferation of human CD4⁺ T cells at low doses (Extended Data Fig. 7h). We have, therefore, developed a novel SPR inhibitor that blocks T cell proliferation and autoimmunity at nanomolar potency and with good oral bioavailability, and this or similar compounds could potentially be used to treat T-cell-mediated autoimmune and allergic diseases.

To investigate whether an increase in GCH1 and BH4 enhances T cell function *in vivo*, we crossed the *Lck-cre* driver line to mice that overexpress a haemagglutinin-tagged human *GCH1* that is inducible by Cre-recombinase¹ to generate *GOE;Lck* animals (Extended Data Fig. 8a). T cell development and homeostasis of peripheral CD4⁺ and CD8⁺ T cells were unaffected in these mice, although in the periphery there was a marked increase in the proportion of effector T cells (Extended Data Fig. 8b–d). Anti-CD3/CD28-stimulated CD4⁺ T cells from the *GOE;Lck* mice had elevated BH4 levels compared with controls (Fig. 4a), and displayed enhanced T cell proliferation upon activation (Fig. 4b). GCH1 overexpression in unstimulated naive T cells did not result in proliferation or any overt spontaneous autoimmunity. To confirm that elevated GCH1 increases the proliferation of activated T cells, we used additional T-cell-specific Cre lines to drive GCH1–HA expression—namely the *Cd4-cre* and tamoxifen-inducible *ERT-cre* lines, both of which show enhanced T cell proliferation and cytokine production (Extended Data Fig. 8e–j). In *GOE;Lck* mice we also observed more inflammatory cells (including T cells) in the ovalbumin-induced allergic inflammation asthma model (Extended Data Fig. 9a), and greater severity in the *in vivo* T-cell-transfer colitis model (Extended Data Fig. 9b, c). Overproduction of BH4 in regulatory T cells did not affect their suppressive function in transfer colitis (Extended Data Fig. 9d). Administration of sepiapterin to anti-CD3/CD28-stimulated CD4⁺ T cells also increased BH4 levels and enhanced the proliferation of stimulated CD4⁺ and CD8⁺ T cells (Fig. 4c and Extended Data Fig. 9e–g). Furthermore, treatment of stimulated CD4⁺ T cells with BH4 itself increased both proliferation and IL-2 secretion (Extended Data Fig. 9h, i), and the proliferative and S-phase-entry defects observed in *Gch1*-ablated T cells were rescued with either sepiapterin (Extended Data

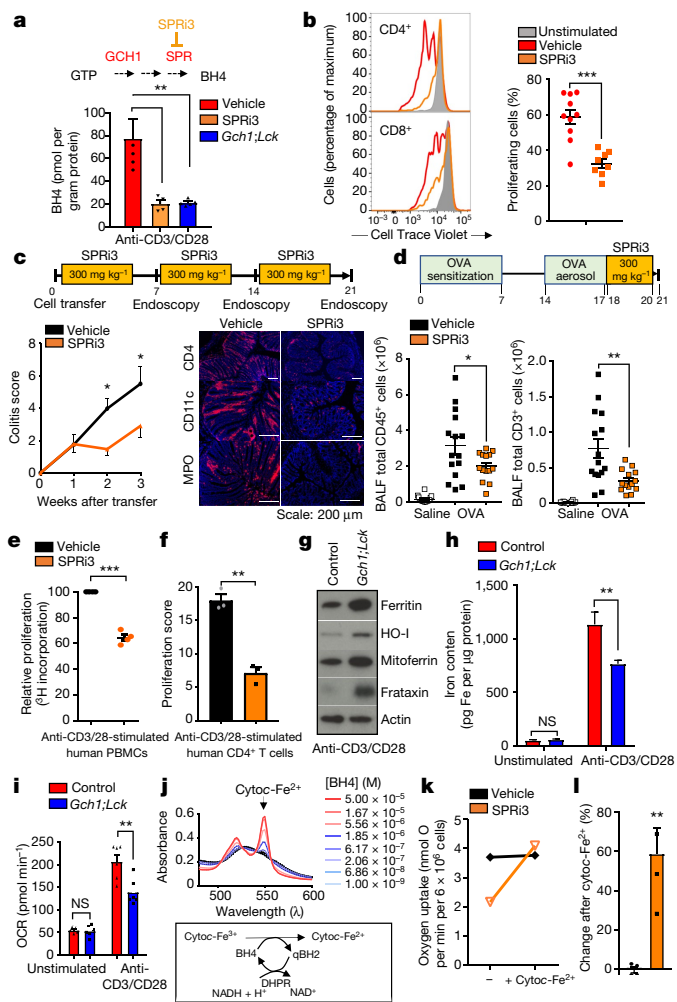


Fig. 2 | Pharmacological inhibition of the BH4 pathway ameliorates T-cell-mediated inflammation. **a**, BH4 production in 24-h-activated control ($n = 5$) and *Gch1*-ablated ($n = 5$) CD4⁺ T cells and in wild-type cells treated with SPRI3 (50 μ M; $n = 4$). Experiment repeated two independent times with similar results. Top, mechanism of action of SPRI3. **b**, Representative 3-day T cell proliferation histogram (right) and quantification (left) of stimulated wild-type T cells (CD4⁺ and CD8⁺) treated with vehicle ($n = 10$) or SPRI3 (50 μ M; $n = 8$). **c**, Colitis model, involving transfer of wild-type CD4⁺ T cells into *Rag1*^{-/-} hosts treated with vehicle or SPRI3 (300 mg kg⁻¹; $n = 8$ each). Right, representative images of intestinal immune infiltration. **d**, Allergic airway inflammatory disease model in control mice treated with SPRI3 (300 mg kg⁻¹; $n = 14$) or vehicle ($n = 15$). **e**, **f**, Proliferation of vehicle- and SPRI3-treated (50 μ M) naive human ($n = 4$ donors) peripheral blood mononuclear cells (PBMCs) (**e**) and purified effector human CD4⁺ T cells (**f**) re-stimulated for 3 days. **g**, Western immunoblot of iron regulators in 24-h-activated peripheral CD4⁺ T cells from control and *Gch1*; *Lck* mice. The experiment was repeated three independent times with similar results. HO-1, haem oxygenase-1. **h**, Total iron content from unstimulated and 24-h-stimulated CD4⁺ T cells from control ($n = 17$) and *Gch1*; *Lck* ($n = 22$) mice. **i**, Oxygen-consumption rate (OCR) in unstimulated and 16-h-stimulated CD4⁺ T cells from control and *Gch1*; *Lck* mice ($n = 6$ each). **j**, Top, dose-dependent reduction of ferro-cytochrome *c* (cytoc-Fe²⁺) to ferro-cytochrome *c* (cytoc-Fe²⁺) by BH4; and bottom, diagram of the reduction pathway. DHPR, dihydropyridine receptor; qBH2, quinonoid dihydrobiopterin. **k**, **l**, Representative oxygen uptake rate in permeabilized, 16-h-stimulated CD4⁺ T cells from vehicle-treated and SPRI3-treated (50 μ M) wild-type cells before and after the addition of cytoc-Fe²⁺ (**k**), and quantification of oxygen rate upon supplementation of cytoc-Fe²⁺ ($n = 4$ independent experiments) (**l**). NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (one-way ANOVA with Dunnett's comparison for **a**; two-tailed Student's *t*-test for **b**, **d**–**f**, **h**, **i**, **l**; two-way ANOVA with Sidak's comparison for **c**). Data are mean \pm s.e.m.

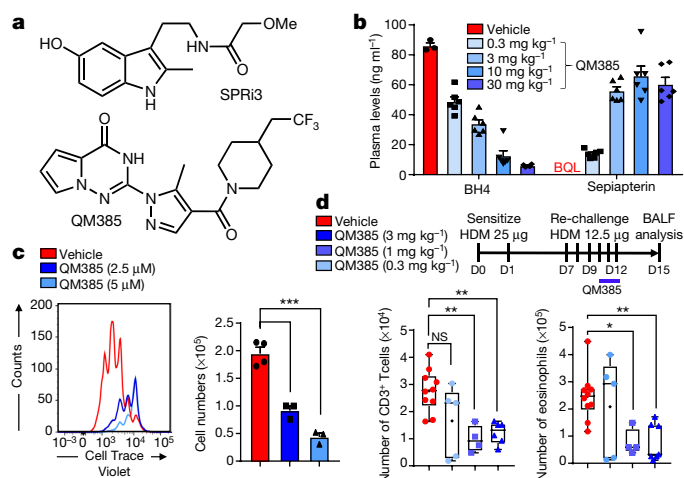


Fig. 3 | Development of an orally available, potent, small-molecule SPR inhibitor to limit BH4 production. **a**, Chemical structure of SPRi3 and QM385. **b**, Dose-dependent reduction in plasma BH4 levels by QM385, and respective dose-dependent increase in plasma sepiapterin levels. BQL, below quantifiable limits (less than 0.3 ng ml⁻¹ for sepiapterin). $n = 6$ mice for each condition. **c**, Left, representative histograms depicting three-day proliferation of stimulated CD4⁺ T cells with vehicle ($n = 4$ mice) and QM385 treatment (2.5 μ M or 5 μ M; $n = 3$ mice each); and right, quantitative analysis of total cell numbers. Experiment repeated two independent times with similar results. **d**, Top, diagram showing the HDM allergy model with dose-response administration of QM385 peritoneally twice a day for three consecutive days as indicated. Bottom, quantification of T cells and eosinophils in BALF. Data are shown as box-and-whisker plots (running from minimal to maximal values), for which individual data points are given. Vehicle, $n = 10$; 0.3 mg kg⁻¹, $n = 5$; 1 mg kg⁻¹, $n = 4$; 3 mg kg⁻¹, $n = 6/7$. Absolute numbers in the BALF are shown. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (one-way ANOVA with Dunnett's multiple comparisons for **d**, **e**). Data in **b**, **c** are mean \pm s.e.m.

Fig. 9e, f) or BH4 (Extended Data Fig. 9j). In activated *Gch1*-ablated T cells, sepiapterin supplementation also restored iron levels, reduced superoxide and increased ATP production (Extended Data Fig. 10a–c), reinforcing that these deficits are due to reduced BH4 levels.

To address whether hyperactivation of the BH4 pathway in T cells promotes anticancer immunity, we orthotopically injected E0771 breast-cancer cells into syngeneic mice to generate mammary tumours²¹. *GOE;Cd4* mice, unlike controls, completely rejected tumour growth (Fig. 4d). Moreover, treatment of mice carrying established E0771-derived mammary tumours with BH4 slowed the growth of the tumours (Fig. 4e). Tumours in BH4-treated mice displayed increased frequencies of activated effector CD4⁺ and CD8⁺ cells among the infiltrating T cells, compared with vehicle-treated mice (Fig. 4f and Extended Data Fig. 10d). BH4 treatment in *Rag2*^{-/-} hosts had no effect on breast-cancer growth, confirming that the effect of BH4 is via effects on the adaptive immune system (Fig. 4g). We validated these results with a second orthotopic model, the TC-1 cancer line (Extended Data Fig. 10e–g). Kynurenine—a tryptophan metabolite—inhibits T cell proliferation²²; xanthurenic acid, a kynurenine metabolite, blocks SPR activity²³. We found that kynurenine treatment inhibits SPR in activated T cells, as shown by increased sepiapterin levels (Extended Data Fig. 10h). Adding kynurenine to T cell cultures also reduced T cell proliferation and increased ROS in activated T cells, both of which were fully restored by addition of BH4 (Fig. 4h, i and Extended Data Fig. 10i, j).

In conclusion, we have revealed that BH4 is required for the effective proliferation of mature T cells in vitro and in vivo, and that this is mechanistically linked to iron metabolism and mitochondrial respiration. Of relevance in this context, nutritional iron deficiency is associated with impaired T cell proliferation and delayed-type hypersensitivity responses, while humoral immunity is largely preserved^{24,25}.

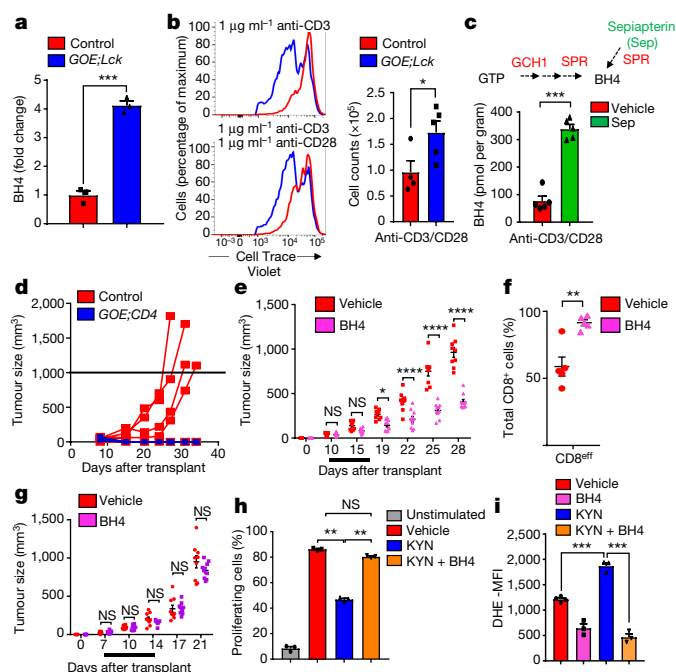


Fig. 4 | Enhanced BH4 production results in enhanced T cell proliferation and anticancer immunity. **a**, Fold change in BH4 levels after 24-h stimulation of CD4⁺ T cells. $n = 3$ individual mice. **b**, Representative histograms after three days of CD4⁺ T cell proliferation, from control ($n = 4$) and *GOE;Lck* ($n = 5$) mice. **c**, BH4 measurements after 24 hours in activated wild-type CD4⁺ T cells treated with vehicle or sepiapterin (Sep; 5 μ M). $n = 5$ individual mice. **d**, Breast-cancer model, involving orthotopic injection of E0771 breast-cancer cells into syngeneic control ($n = 6$) and *GOE;Lck* ($n = 7$) mice. **e**, Effect of BH4 supplementation on cancer growth in the E0771 model. Supplementation with BH4 ($n = 10$ mice) or vehicle ($n = 9$ mice) was carried out for seven days as indicated (black line). **f**, Quantification of intratumoral effector CD8⁺ T cells (CD44⁺CD62L^{low}) assayed from E0771 tumours on day 28 of vehicle treatment ($n = 5$ mice) or BH4 treatment ($n = 5$ mice). **g**, Effect of BH4 supplementation on cancer growth in *Rag2*^{-/-} female hosts. BH4 and vehicle supplementation ($n = 9$ mice each) was carried out for seven days as indicated (black line). **h**, Quantification of proliferation of stimulated CD4⁺ T cells treated with kynurenine (KYN; 50 μ M) or BH4 (10 μ M). $n = 3$ samples for each condition. Experiment was repeated two independent times with similar results. **i**, Quantification of the mean fluorescent intensity (MFI) of dihydroethidium (DHE, a superoxide ROS indicator) in stimulated wild-type CD4⁺ T cells treated with vehicle, KYN (50 μ M), BH4 (10 μ M) or KYN (50 μ M) plus BH4 (10 μ M) for 20 hours. $n = 3$ samples for each condition. The experiment was repeated two independent times with similar results. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$ (two-tailed Student's *t*-test for **a**, **b**, **c**, **f**; two-way ANOVA with Sidak's comparison for **e**, **g**; one-way ANOVA with Tukey's comparison for **h**, **i**). Data are mean \pm s.e.m.

Iron-deficiency anaemia is also associated with an increased incidence of cancer^{26,27}. Notably, we have further found that BH4 is required for T-cell-driven autoimmunity and allergic inflammation and that its inhibition by kynurenine links the immunosuppressive tumour environment to impaired T cell function. Moreover, increasing BH4 levels can overcome this inhibition to enhance immunity and inhibit tumour growth. Therefore, blocking the synthesis of BH4 could be a viable way to abrogate proinflammatory auto-aggressive T cells in T-cell-driven pathological diseases, whereas its elevation could be a novel way to enhance antitumour immunity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0701-2>.

Received: 18 December 2017; Accepted: 20 September 2018;
Published online 7 November 2018.

- Latremoliere, A. et al. Reduction of neuropathic and inflammatory pain through inhibition of the tetrahydrobiopterin pathway. *Neuron* **86**, 1393–1406 (2015).
- Werner, E. R., Blau, N. & Thöny, B. Tetrahydrobiopterin: biochemistry and pathophysiology. *Biochem. J.* **438**, 397–414 (2011).
- Chen, W. et al. Role of increased guanosine triphosphate cyclohydrolase-1 expression and tetrahydrobiopterin levels upon T cell activation. *J. Biol. Chem.* **286**, 13846–13851 (2011).
- Ziegler, I. et al. Control of tetrahydrobiopterin synthesis in T lymphocytes by synergistic action of interferon- γ and interleukin-2. *J. Biol. Chem.* **265**, 17026–17030 (1990).
- Chuaipichai, S. et al. Cell-autonomous role of endothelial GTP cyclohydrolase 1 and tetrahydrobiopterin in blood pressure regulation. *Hypertension* **64**, 530–540 (2014).
- Eberl, G. & Littman, D. R. Thymic origin of intestinal $\alpha\beta$ T cells revealed by fate mapping of ROR γ^t cells. *Science* **305**, 248–251 (2004).
- Hobeika, E. et al. Testing gene function early in the B cell lineage in mb1-cre mice. *Proc. Natl Acad. Sci. USA* **103**, 13789–13794 (2006).
- Śledzińska, A. et al. TGF- β signalling is required for CD4 $^+$ T cell homeostasis but dispensable for regulatory T cell function. *PLoS Biol.* **11**, e1001674 (2013).
- Talbot, S. et al. Silencing nociceptor neurons reduces allergic airway inflammation. *Neuron* **87**, 341–354 (2015).
- Haworth, O., Cernadas, M., Yang, R., Serhan, C. N. & Levy, B. D. Resolvin E1 regulates interleukin 23, interferon- γ and lipoxin A4 to promote the resolution of allergic airway inflammation. *Nat. Immunol.* **9**, 873–879 (2008).
- Martin, S. F. et al. Toll-like receptor and IL-12 signaling control susceptibility to contact hypersensitivity. *J. Exp. Med.* **205**, 2151–2162 (2008).
- Rangachari, M. & Kuchroo, V. K. Using EAE to better understand principles of immune function and autoimmune pathology. *J. Autoimmun.* **45**, 31–39 (2013).
- Rangachari, M. et al. Bat3 promotes T cell responses and autoimmunity by repressing Tim-3-mediated cell death and exhaustion. *Nat. Med.* **18**, 1394–1400 (2012).
- Nar, H. et al. Active site topology and reaction mechanism of GTP cyclohydrolase I. *Proc. Natl Acad. Sci. USA* **92**, 12120–12125 (1995).
- Nar, H. et al. Atomic structure of GTP cyclohydrolase I. *Structure* **3**, 459–466 (1995).
- Volani, C. et al. Dietary iron loading negatively affects liver mitochondrial function. *Metalomics* **9**, 1634–1644 (2017).
- Archer, M. C., Vonderschmitt, D. J. & Scrimgeour, K. G. Mechanism of oxidation of tetrahydropterins. *Can. J. Biochem.* **50**, 1174–1182 (1972).
- Eberlein, G., Bruice, T. C., Lazarus, R. A., Henrie, R. & Benkovic, S. J. The interconversion of the 5,6,7,8-tetrahydro-, 7,8-dihydro-, and radical forms of 6,6,7,7-tetramethyldihydropterin. A model for the biopterin center of aromatic amino acid mixed function oxidases. *J. Am. Chem. Soc.* **106**, 7916–7924 (1984).
- Capeillere-Blandin, C., Mathieu, D. & Mansuy, D. Reduction of ferric haemoproteins by tetrahydropterins: a kinetic study. *Biochem. J.* **392**, 583–587 (2005).
- Hondowicz, B. D. et al. Interleukin-2-dependent allergen-specific tissue-resident memory cells drive asthma. *Immunity* **44**, 155–166 (2016).
- Ewens, A., Mihich, E. & Ehrke, M. J. Distant metastasis from subcutaneously grown E0771 medullary breast adenocarcinoma. *Anticancer Res.* **25**, 3905–3915 (2005).
- Curti, A. et al. Indoleamine 2,3-dioxygenase-expressing leukemic dendritic cells impair a leukemia-specific immune response by inducing potent T regulatory cells. *Haematologica* **95**, 2022–2030 (2010).
- Haruki, H., Hovius, R., Pedersen, M. G. & Johnsson, K. Tetrahydrobiopterin biosynthesis as a potential target of the kynurenine pathway metabolite xanthurenic acid. *J. Biol. Chem.* **291**, 652–657 (2016).
- Oppenheimer, S. J. Iron and its relation to immunity and infectious disease. *J. Nutr.* **131**, 616S–633S (2001).
- Cassat, J. E. & Skaar, E. P. Iron in infection and immunity. *Cell Host Microbe* **13**, 509–519 (2013).
- Liu, C.-J., Chen, K.-W., Hu, Y.-W., Hong, Y.-C., Huang, Y.-C., Chiou, T.-J. & Tzeng, C.-H. Chronic iron deficiency anemia and cancer risk. *Blood* **120**, 5172 (2012).
- Hung, N. et al. Risk of cancer in patients with iron deficiency anemia: a nationwide population-based study. *PLoS ONE* **10**, e0119647 (2015); correction <https://doi.org/10.1371/journal.pone.0125951> (2015).

Acknowledgements We thank all members of our laboratories for helpful discussions and Life Science Editors for editorial support. We thank Shanghai ChemPartners for running the drug metabolism and pharmacokinetic assays associated with QM385. J.M.P. is supported by grants from IMBA, the Austrian Ministry of Sciences and the Austrian Academy of Sciences, and the T. Von Zastrow Foundation as well as a European Research Council (ERC) Advanced Grant and an Era of Hope Innovator award. C.J.W. is supported by a National Institutes of Health (NIH) R35 grant (NS105076). We also acknowledge the Christian Doppler Laboratory for Iron Metabolism and Anemia Research as a funding body for our research (M.S. and G.W.). M.R. is supported by EMD Serono, Canada, and a MS Network Transitional Career Development Award.

Reviewer information Nature thanks R.S. Johnson, L. O'Neill and N. Restifo for their contribution to the peer review of this work.

Author contributions S.J.F.C., together with C.J.W. and J.M.P., conceived and designed the study. All experiments were performed by S.J.F.C. with the following exceptions: A.W. and A.J. performed mitochondrial respiration analyses; S. Reissig performed colonoscopy grading; S.T., C.S. and B.L.T. carried out the asthma model; C.S. and B.L.T. carried out the HDM model; Y.P. performed the iron-reduction experiment; M.S.L., G.L. and G.W. carried out assays for human T cell proliferation; M.S. performed the iron measurements; T.K. performed in vitro thymocyte differentiation experiments; M.N. performed microarray analysis; E.M., B.L.T. and D.d.L.S. performed biopterin and sepiapterin measurements; M.R., M.K., D.H., M.T., L.T., D.C., S. Rao, M.P. and M.A. helped with the cancer studies; L.B., N.A., A. Latremoliere and M.C. helped with compound dosing and discussions of BH4 biology; M.T. and S.Z. performed QM385 pharmacokinetic analysis. S.J.F.C., C.J.W. and J.M.P. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0701-2>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0701-2>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to C.J.W. or J.M.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mice. Mice expressing enhanced GFP (eGFP) under the *Gch1* promoter were used to label cells that upregulate *Gch1* after T cell activation. Mice with a *cre*-dependent *GCH1*-HA overexpression cassette to induce BH4 overproduction, and *Gch1* floxed mice in which BH4 production is prevented, have previously been reported^{1,5}. For both gain- and loss-of-function experiments, we bred *GCH1*-HA and *Gch1* floxed mice to the T-cell-specific lines *LCK-cre*²⁸, *Cd4-cre*²⁹ or *RORgammat-cre*, the ubiquitous tamoxifen-inducible *Rosa26-cre*^{ERT2} line³⁰, and the B-cell-specific line *MB1-cre*. All animal experiments were approved by the Austrian Animal Care and Use Committee.

Compounds. Sepiapterin (Sep; 11,225), tetrahydrobiopterin (BH4, 11,212) were purchased from Schircks Labs. For *in vitro* use, both sepiapterin and BH4 were dissolved in dimethylsulfoxide (DMSO) to a stock concentration of 10 mM. SPRI3 has previously been developed and was used as instructed¹. For T cell assays, sepiapterin was used at a concentration of 5 μ M, BH4 at a concentration of 10 μ M and SPRI3 at a concentration of 50 μ M unless otherwise stated in the figure legends. For *in vivo* use, BH4 was reconstituted in sterile saline under argon gas. Kynurenine (K8625) and NAC (A9165) were purchased from Sigma.

Determination of BH4 levels. Levels of BH4 (tetrahydrobiopterin) and oxidized biopterins (BH2 and biopterin) were determined by high-performance liquid chromatography (HPLC) followed by electrochemical and fluorescent detection, respectively, following an established protocol³¹. Cell pellets were freeze-thawed in ice-cold resuspension buffer (50 mM phosphate-buffered saline (PBS), 1 mM dithioerythritol, 1 mM EDTA, pH 7.4). After centrifugation at 13,200 r.p.m. for 10 min at 4°C, supernatant was removed and ice-cold acid precipitation buffer (1 M phosphoric acid, 2 M trichloroacetic acid and 1 mM dithioerythritol) was added. Following centrifugation at 13,200 r.p.m. for 10 min at 4°C, the supernatant was removed and injected onto the HPLC system. Quantification of BH4 and oxidized biopterins was obtained by comparison with external standards and normalized to protein concentration, determined by the bicinchoninic acid (BCA) protein assay (Pierce).

Determination of sepiapterin levels by HPLC. Supernatant samples were precipitated by the addition of one volume (1/1, v/v) of 5% trichloroacetic acid (TCA) plus 6.5 mM dithiothreitol (DTT). Afterwards, samples were centrifuged (10,000 g for 10 min at 4°C) and 20 μ l was analysed. HPLC analysis of sepiapterin was done using a Beckman System Gold (Beckman Instruments) by using a Waters Atlantis dC-18, 5- μ m RP column (4.6 mm \times 250 mm; temperature 35°C), with a flow rate set at 0.5 ml min⁻¹ and isocratic elution of mobile phase (92% phosphate buffer (15 mM); 8% acetonitrile (90%), pH 6.4). Identification and quantification of sepiapterin was done using a multiwavelength fluorescence detector (excitation wavelength 425 nm, emission wavelength 530 nm, module 2,475; Waters) and expressed as nM of sepiapterin.

Lymphocyte proliferation. T cells were purified from spleens and lymph nodes of mice using microbeads (CD4⁺; CD8⁺, naive CD4⁺, Miltenyi Biotec). We coated 96 U-shaped plates with anti-CD3 (4 μ g ml⁻¹; Biolegend), with or without anti-CD28 (2 μ g ml⁻¹; Biolegend) at the indicated concentrations (unless otherwise stated in the figure legends) in PBS for 3 h at 37°C. T cells were then plated at 10⁵ cells per well in Iscove's modified Dulbecco medium (IMDM) plus penicillin streptomycin plus L-glycine plus 10% fetal calf serum (FCS). Beta-mercaptoethanol was omitted. Phorbol myristate acetate (50 ng ml⁻¹) and ionomycin (50 ng ml⁻¹) were also used to stimulate purified T cells for 24 h. Purified and activated T cells were cultured for 24 h; the expression of activation markers (CD62L, CD25, CD44 and CD69) was analysed using flow cytometry; and the supernatant was collected for measurement of IL-2 and IFN γ concentrations using ELISA kits (Biolegend). Purified T cells were also stained with the Cell Violet Trace Proliferation Kit (Invitrogen) and cultured for three days; proliferation was assayed by flow cytometry using viable cells (4',6'-diamidino-2-phenylindole (DAPI)-negative). In addition, purified T cells were cultured with purified splenic dendritic cells and soluble anti-CD3 antibody (1 μ g ml⁻¹) for three days. To analyse the expression of inducible (i)NOS, we stimulated purified CD4⁺ T cells, assayed fixed and permeabilized cells at various time points, and stained them for iNOS levels. B cells were purified using microbeads (CD19⁺; Miltenyi Biotec) from the spleen, loaded with cell tracer, stimulated with lipopolysaccharide (LPS; 1 μ g ml⁻¹) and analysed for proliferation as described above. For the class-switch recombination experiment, CD43⁻ B cells were isolated from spleens by magnetic-activated cell sorting (MACS; Miltenyi Biotec) and stimulated for five days with LPS (20 μ g ml⁻¹) to induce switching to IgG3 expression. Percentages of switched B lymphocytes were assessed by flow cytometry.

EdU staining. The cell-cycle status of T cells was assessed using the Click-iT EdU flow cytometry cell proliferation assay (Invitrogen). In brief, purified CD4⁺ T cells were activated with anti-CD3 (4 μ g ml⁻¹) and anti-CD28 (2 μ g ml⁻¹) as described above. EdU was pulsed into the wells for 4 h after 16 h of stimulation. The cells were prepared and stained with EdU as per the manufacturer's instructions.

Mitochondrial respiration and metabolomics. Mitochondrial respiratory parameters were measured with high-resolution respirometry (Oxygraph-2k, Oroboros

Instruments)³². Routine respiration was measured by incubating cells in a buffer containing 110 mM sucrose, 60 mM K-lactobionate, 20 mM K-HEPES, 10 mM KH₂PO₄, 3 mM MgCl₂, 0.5 mM egtazic acid (EGTA) and 1 g l⁻¹ fatty-acid-free bovine serum albumin at 37°C (pH 7.2). Total capacity was induced by titration of carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (Sigma Aldrich) in steps of 0.5 μ M. To assess complex-I- and complex-II-linked respiration, cells were permeabilized with digitonin (8 μ M). Complex-I-linked state 3 respiration was induced by adding 5 mM glutamate/5 mM malate and 1 mM adenosine diphosphate (ADP). Complex-II-linked state 3 respiration was induced with 10 mM succinate after adding the complex I inhibitor rotenone (1 ng ml⁻¹). To restore respiratory function in activated CD4⁺ T cells, cells were permeabilized with digitonin (12 μ M) and exogenous reduced cytochrome *c* (2.5 μ M; Abcam, b140219) was added. Respiration rates were obtained by calculating the negative time derivative of the measured oxygen concentration. Oxygen-consumption rates were measured using Seahorse technology. To measure ATP, purified T cells were either left unstimulated or stimulated with plate-bound anti-CD3 (4 μ g ml⁻¹) and anti-CD28 (2 μ g ml⁻¹) for the times indicated in the figures. ATP was measured using the CellTiter-Glo Luminescent cell viability assay (Promega). To determine ROS levels, purified T cells were activated with anti-CD3 plate-bound anti-CD3 (4 μ g ml⁻¹) and anti-CD28 (2 μ g ml⁻¹) for 10 h. Cells were washed once with Hank's balanced salt solution (HBSS) and stained in 10 μ M DHE (Invitrogen) for 30 min at 37°C. Cells were washed twice with HBSS and assayed by flow cytometry. Profiling of biogenic amines by hydrophilic interaction liquid chromatography/quadrupole time-of-flight mass spectrometry was performed on cell pellets and supernatants from unstimulated and TCR-stimulated purified T cells by the West Coast Metabolomics Center (UC Davis). For NO₂ measurements, we used the Griess reagent system (TB229, Promega). Purified T cells were stimulated with anti-CD3 and anti-CD28 antibodies as described above, and the supernatant was collected at various time points for nitrite measurements. Peritoneal macrophages stimulated with LPS (100 ng ml⁻¹) for 24 h were used as a positive control.

Flow cytometry. Antibody labelling of cells was carried out in FACS staining buffer (PBS supplemented with 2% FCS and 2 mM EDTA) on ice for 30 min after blocking Fc receptors. See Supplementary Table 4 for a list of antibodies used in this study. Cells were recorded on an LSR II flow cytometer (BD Biosciences), and data were analysed using FlowJo v10.0.6 software (Tree Star). Absolute splenocyte and thymus numbers were determined by counting total cells with a CASY1 counter, and subsequent calculation of T cell and B cell numbers was based on ratios from FACS experiments.

Protein blotting. Protein blotting was carried out using standard protocols. Blots were blocked for 1 h with 5% bovine serum albumin (BSA) in TBST (1 \times Tris-buffered saline (TBS) and 0.1% Tween-20) and were then incubated overnight at 4°C with primary antibodies (see Supplementary Table 4), diluted in 5% BSA in TBST (1/1,000 dilution). Blots were washed three times in TBST for 15 min and were then incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (1/2,500 dilution; GE Healthcare, NA9340V) for 45 min at room temperature, washed three times in TBST for 15 min and visualized using enhanced chemiluminescence (ECL Plus, Pierce, 1896327).

OP9-DL1 co-cultures. OP9 bone marrow stromal cells expressing the Notch ligand DL-1 (OP9-DL1; provided by J. C. Zúñiga-Pflücker, University of Toronto) were maintained as described³³. We plated 10⁴ OP9-DL1 per well in 48-well plates 4–12 h before the start of thymocyte cultures. DN3a thymocytes were sorted as TCR β ⁺TCR γ δ ⁻CD4⁻CD8a⁻CD28⁻CD25^{high}CD44^{low} cells using a BD FACS Aria sorter. Cell Trace Violet labelling of the sorted cells was performed in 1 μ M Cell Trace Violet solution in PBS containing 0.1% BSA for 7 min at 37°C. Cells were washed with medium containing 20% FCS. Thymocytes were then plated on the OP9-DL1 monolayers in the presence of 5 ng ml⁻¹ Flt3L (a tyrosine-kinase-3 inhibitor). Co-cultures were performed in α -minimal essential medium (α MEM) supplemented with 10 mM HEPES (pH 7.5), 1 mM sodium pyruvate, 100 units per ml penicillin, 0.1 mg ml⁻¹ streptomycin, and 20% heat-inactivated FBS.

Adoptive transfer model of colitis. We injected 5 \times 10⁵ MACS-purified naive CD4⁺CD62L⁺ T cells from control and *GCH1*;*Lck* mice intraperitoneally into 6- to 8-week-old *Rag1*^{-/-} mice. After the cell transfer, *Rag1*^{-/-} recipients were weighed weekly and monitored by mini-endoscopy. For monitoring of colitis activity, we used a high-resolution video endoscopic system (Karl Storz). To determine colitis activity, we anaesthetized mice by injecting intraperitoneally a mixture of ketamine (Ketavest 100 mg ml⁻¹; Pfizer) and xylazine (Rompun 2%; Bayer Healthcare) and monitored them by mini-endoscopy at the indicated time points. Endoscopic scoring of five parameters (translucency, granularity, fibrin, vascularity and stool) was performed (Supplementary Table 5)³⁴. For histological analysis, colonic cross-sections were stained with haematoxylin and eosin (H&E). Immunofluorescence of cryo-sections was performed using the tyramide signal amplification (TSA) Cy3 system (PerkinElmer) and a fluorescence microscope (IX70; Olympus) using primary antibodies against F4/80, MPO, CD3, CD4 and CD11c. In brief, cryo-sections were fixed in ice-cold acetone for 10 min, and then

incubated sequentially with methanol, avidin/biotin (Vector Laboratories) and protein-blocking reagent (DAKO) to eliminate unspecific background staining. Slides were then incubated overnight with primary antibodies specific for the respective antigen. Subsequently, the slides were incubated for 30 min at room temperature with biotinylated secondary antibodies (Dianova). All samples were finally treated with streptavidin–HRP and stained with tyramide (Cy3) according to the manufacturer's instructions (Perkin Elmer). Before examination, nuclei were counterstained with Hoechst 3342 (Invitrogen). For experiments involving transfer of regulatory T cells³⁵, 500,000 conventional T cells (CD4⁺CD25[−]CD45RB^{high}) and 500,000 regulatory T cells (CD4⁺CD25⁺CD45RB^{low}) were transferred intraperitoneally into *Rag2*^{−/−} hosts. For *GOE*; *Cd4* CD4⁺ T cells, 150,000 cells were transferred. Body weights were monitored over the course of the experiment³⁵.

OVA immunization and airway hyperresponsiveness. For the OVA immunization study, immunization was performed using 100 µg OVA per mouse in 200 µl alum intraperitoneally. Blood was collected from the tail vein 14 days after injection to check IgG and IgM titres. Three weeks later, a further intraperitoneal injection was carried out and again blood collected two weeks later to measure the challenge responses. For measurements of lung function, deeply anaesthetized mice (pentobarbital (60 mg kg^{−1}) underwent a tracheotomy with a 20G sterile catheter. A computer-based analysis of airway hyperresponsiveness was then performed using a Flexivent (SCIREQ) apparatus¹⁰. Mice were ventilated at a tidal volume of 9 ml kg^{−1} with a frequency of 150 b.p.m.; positive end-expiratory pressure was set at 2 cm H₂O. Lung resistance and elastance of the respiratory system was determined in response to in-line aerosolized methacholine challenges (0, 1, 3, 10, 30 and 100 mg ml^{−1}). Methacholine was dissolved in sterile PBS. The mean elastance and resistance of ten measurements by doses was calculated. For bronchoalveolar lavage (BAL) on day 21, mice were anaesthetized following an intraperitoneal injection of urethane (200 µl; 35%) and a 20G sterile catheter inserted longitudinally into the trachea. We injected 2 ml of ice-cold PBS containing protease inhibitors (Roche) into the lungs, collected and stored on ice. BAL fluid underwent a 400g centrifugation (15 min; 4 °C), the supernatant was discarded and cells were resuspended in 200 µl BAL fluid¹⁰. Bronchoalveolar lavage fluid (BALF) cells were resuspended in FACS buffer (PBS, 2% FCS, EDTA), and incubated with Fc block (0.5 mg ml^{−1}; 10 min; BD Biosciences). Cells were then stained with monoclonal antibodies (FITC-conjugated anti-mouse CD45, BD Biosciences, catalogue number 553,079; phycoerythrin (PE)-conjugated anti-mouse Syglec-F, BD Biosciences, catalogue number 552,126; allophycocyanin (APC)-conjugated anti-mouse GR-1, eBiosciences, catalogue number 17-5931-81; PE-Cy7-conjugated anti-mouse CD3ε, catalogue number 25-0031-81; peridinin chlorophyll protein complex (PerCP)-conjugated anti-mouse F4/80, BioLegend, catalogue number 123,125; 45 min, 4 °C on ice) before data acquisition on a FACS Canto II (BD Biosciences). A leukocyte differential count was performed during flow-cytometry analysis of cells expressing the common leukocyte antigen CD45 (BD Pharmingen; catalogue number 553,079). Specific cell populations were identified as follows: macrophages as F4/80^{high}Ly6g[−], eosinophils as F4/80^{int}Ly6g^{low}SiglecF^{high}, neutrophils as F4/80^{low}Ly6g^{high}SiglecF[−], and T lymphocytes as F4/80[−]Ly6g[−]CD3⁺. Total BAL cell counts were performed using a standard haemocytometer, with absolute cell numbers calculated as total BAL cell number multiplied by the percentage of cell subpopulation as determined by FACS¹⁰.

HDM allergy model. For HDM-induced lung inflammation, C57BL/6 animals (female, 6–12 weeks old) were sensitized for two consecutive days with 25 µg HDM extract (*Dermatophagoides pteronyssinus*, Greer Laboratories, XPB82D382.5) intranasally. Six days after the last sensitizing dose, mice were challenged with 12.5 µg of HDM for five consecutive days, with 3 mg kg^{−1} QM-760 (in 1% Tween-80 and 0.5% sodium carboxymethyl cellulose; Sigma–Aldrich) administered by oral gavage, twice daily on days 3–5 during the challenge phase. BALF was removed and analysed for the following immune-cell subsets three days after the last challenge—T cells: CD45⁺, Thy1⁺, CD3⁺, CD11b[−], Siglec-F[−], Ly6C/G[−]; eosinophils: CD45⁺, Thy1⁺, CD11b⁺, Siglec-F⁺, CD3[−], Ly6C/G[−].

Skin hypersensitivity. The skin-contact hypersensitivity model was performed as described¹¹. In brief, to induce contact hypersensitivity, mice were sensitized on day zero by applying 100 µl of 7% 2,4,6-trinitrochlorobenzene (TNCB; Sigma)/acetone or acetone alone as a vehicle control on the shaved abdomen. On day five, mice were challenged on the dorsum of both ears with 20 µl of 1% TNCB/acetone. Ear thickness was measured immediately before and 24 h after the challenge.

Experimental autoimmune encephalitis. Experimental autoimmune encephalitis (EAE) was induced in control and *Gch1*; *Lck* mice by immunization with an emulsion of 100 µg myelin oligodendrocyte glycoprotein (MOG)_{35–55} in complete Freund's adjuvant (CFA), supplemented with 5 mg ml^{−1} *Mycobacterium tuberculosis* (Difco). We injected 100 µl MOG/CFA subcutaneously above the inguinal lymph node on both sides of the mouse. We then injected 200 µl pertussis toxin/PBS (50 ng µl^{−1}; List Biological Labs) intraperitoneally per mouse on days zero and one. Scoring for EAE was performed as described over the course of 45 days³⁶.

Orthotopic cancer models. E0771 cells were orthotopically injected into syngeneic control and *GOE*; *Lck* mice as described²¹. In brief, cells were collected for injection into mice by trypsin digestion for 5 min, washed in HBSS, counted, diluted in this salt solution and orthotopically injected into the fat pad of the fourth mammary gland (2.5 × 10⁵ cells per 200 µl per mouse). BH4 administration was delivered intraperitoneally (100 mg kg^{−1}) after tumours were palpable (day 10) and treatment was continued for seven days. Tumours were measured using digital calipers; the size of the tumour was expressed as length (mm) by width (mm) by height (mm) equals tumour size (mm³). The tumour cell line TC-1 was derived from primary lung epithelial cells of C57BL/6 mice. The cells were immortalized with the amphotropic retrovirus vector LXSN16E6E7 and subsequently transformed with the pVEJB plasmid expressing the activated human *c-H-RAS* oncogene³⁷. This cell line was treated and injected into wild-type and *Rag2*^{−/−} mice as described above. After collagenase/dispase digestion of the tumours, intratumoral effector CD4⁺ and CD8⁺ T cells were characterized by flow cytometry (CD62L^{low}, CD44^{high}).

Microarray analysis. Purified CD4⁺ T cells from control and *Gch1*; *Lck* mice were stimulated with plate-bound anti-CD3 (4 µg ml^{−1}) and anti-CD28 (2 µg ml^{−1}) antibodies for 16 h, and total RNA was extracted by sequential Qiazol extraction and purification through the RNeasy micro kit with on-column genomic DNA digestion (Qiagen). RNA quality was determined by an Agilent 2100 Bioanalyzer using the RNA Pico Chip (Agilent). RNA was amplified into complementary DNA using the Ambion wild-type expression kit for whole-transcript expression arrays, with poly-A controls from the Affymetrix Genechip Eukaryotic Poly-A RNA control kit. Images from Agilent arrays were processed using Agilent Feature Extraction Software 10.7.3.1. Raw intensity data were processed in R v3.4.0 using limma v3.34.3, applying normexp background calculation, lowess within-array and Aquantile between-array normalization methods. The normalized values were used to calculate log₂-transformed Cy5/Cy3 ratios. Differential-expression analysis was performed by fitting a linear model to the normalized data and computing empirical Bayes test statistics in limma, accommodating a mean-variance trend³⁸. The false-discovery rate was controlled by Benjamini–Hochberg adjustment. The data discussed herein have been deposited in the National Center for Biotechnology Information (NCBI)'s Gene Expression Omnibus³⁹ and are accessible through GEO accession number GSE108101 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108101>).

Ferric/ferrous reduction. The enzymatic activity of BH4 was assayed as described⁴⁰. The enzymatic conversion of qBH2 to BH4 was followed by the reduction of ferricytochrome *c* to ferrocyclochrome *c* by BH4. Ferricytochrome *c* reduction was determined by reading the increasing ferrocyclochrome *c* absorbance signal at 550 nm. The experiment was run for 40 min at pH 7.4 and recorded at 10-s intervals in 200 µl buffer containing 50 µM ferricytochrome *c*, 1 µM 6-methyltetrahydropterin (6MPH4), 20 nM DHPR, 50 µM NADH and selected inhibitors. A control lacking DHPR was run in parallel to assess the rate of non-enzymatic reduction of qBH2 by NADH. The extinction coefficients used for ferrocyclochrome *c* and NADH are respectively 29,500 (reduced, 550 nm, H₂O) and 6,220 (340 nm, H₂O) (in l mol^{−1} cm^{−1}). Generally, 50 µM of ferrocyclochrome *c* and ferricytochrome *c* in buffer were measured in isolated wells to assess the completion of the reaction.

Iron measurements. Total iron content was measured as described⁴¹. In brief, intracellular iron measurements were carried out by using a PerkinElmer Analyst 800 equipped with a transversely heated graphite atomizer. A Zeeman-effect background correction was realized by a 0.8-T magnetic field, oriented longitudinally with respect to the optical path. A PerkinElmer Lumina single-element iron hollow cathode lamp was driven at a constant current of 30 mA after proper equilibration (that is, for 20 min or more). For the absorption measurements, the 248.3-nm line (spectral bandwidth 0.7 nm) was chosen. FACS-purified naive CD4⁺ T cells from control and *Gch1*; *Lck* mice were left untreated or stimulated (anti-CD3 and anti-CD28) for 12 h. The cells were then pelleted and frozen at −80 °C. Next, the samples were suspended in 200 µl of a 0.1% (v/v) solution of nitric acid (Rotipuran Supra, 69%, Carl Roth GmbH) in high-purity water (Milli-Q, Merck-Millipore) by extended periods (that is, for 30 min or more) of vortexing and ultrasonication at 30–40 kHz. After an initial estimation of the sample's iron quantity, a five-point linear calibration was established in the range between 0 (that is, less than 0.004 µM) and 0.106 µM. The calibration standards were prepared by diluting a 0.1 M standard stock solution of (NH₄)₂Fe(SO₄)₂ (Merck-Millipore) with a 0.1% (v/v) aqueous solution of nitric acid (vide supra). The absence of detectable iron (that is, less than 0.004 µM) in the dilution agent, as well as in the sample cups, and the glassware was verified throughout the analyses. A linear fit of the 15 data points ($k = 0.978$, $d = 0.006$ µM) yielded a coefficient of determination of 0.992 where k is the slope of the linear fit and d is the axis intercept on the y-axis. Samples with iron concentrations exceeding the calibration range (that is, 0.106 µM or more) were diluted appropriately. The blank solution, the calibration standards and the samples were supplied to the atomizer in randomized fashion as triplicates, using a PerkinElmer AS-800 autosampler with an injection volume of 20 µl.

The solvent was evaporated by a slow temperature gradient to 130 °C; ashing took place at a maximum temperature of 1,000 °C; and the atomization profile was read at 2,000 °C. The graphite tube, which was protected against oxidation by high-purity argon (99.999%; Messer Austria GmbH), was cleaned out after each analysis at 2,450 °C. The integrity of each analysis was verified by a visual inspection of the respective time-dependent atomization profile.

Human T-cell-proliferation assays. Proliferation of PBMCs, obtained from healthy blood donors, was assessed following cell exposure to Dynabeads human T-activator CD3/CD28 (bead/cell ratio 1/2) and IL-2 (30 international units per ml), in the absence or presence of vehicle (DMSO) or SPRI3 (50 µM). PBMCs were resuspended in RPMI 1640 medium supplemented with 2 mM L-glutamine, 100 units per ml penicillin, 100 mg ml⁻¹ streptomycin, 1% non-essential amino acids and 10% FBS, seeded at 2.5×10^5 cells per well and cultured for five days. For the last 18 h of culture, cells were pulsed with 0.25 mCi per well ³H-thymidine. Incorporated thymidine was measured by liquid scintillation spectroscopy. In addition, we also determined the proliferation of alloreactive human T cells. PBMCs from a healthy donor were stimulated with M21 tumour cells. Alloreactive T cells (based on MHC mismatch) were cultured for two weeks. Afterwards, effector CD4⁺ T cells were sorted (with regulatory T cells excluded), labelled with carboxyfluorescein succinimidyl ester (CFSE), and stimulated with anti-CD3 and anti-CD28 antibodies for five days, supplemented with either DMSO or SPRI3 (50 µM). For QM385 studies, PBMCs from two donors were stimulated with plate-bound anti-CD3 and anti-CD28 antibodies (1 µg ml⁻¹ each). On day three of stimulation, the number of CD4⁺ T cells was counted by FACS. PBMCs were isolated from healthy subjects (from the Blood Donor Center at the Children's Hospital Boston). Human studies received Institutional Review Board (IRB) approval (number 2011P000202) from the Beth Israel Deaconess Medical Center Ethics Committee, and written consent was obtained from all study participants before inclusion in the study.

QM385 compound analysis. For the SNAP (soluble NSF attachment proteins)-based competition time-resolved FRET (TR-FRET) assay, purified SNAP-SPR and sulfasalazine-SNAP-meGFP were labelled with a twofold excess of benzylguanine/terbium cryptate conjugate (K2-benzylamide-BG; Cisbio) or benzylguanine/sulfasalazine (BG-SSZ), respectively, and purified with NAP5 columns (GE Healthcare) to remove excess labelling reagents. The final reaction mixture contained 2.0 nM terbium-SNAP-SPR, 70 nM SSZ-SNAP-meGFP, 10 µM NADPH and 10 µM NADP⁺ in buffer A (50 mM HEPES-NaOH pH 7.4, 0.15 M NaCl, 0.5 µg ml⁻¹ BSA, 0.05% Triton X-100, 1 mM DTT). Signal was measured after 3 h of incubation with varying concentrations of QM385, using Infinite F500 (TECAN). The excitation wavelength was 320 nm and emission wavelengths were 485 nm (K2-benzylamide-BG; Cisbio) and 520 nm (BG-SSZ), respectively. For BH4 measurements, 50,000 human PBMCs were plated in 96-well plates coated with 1 µg ml⁻¹ human anti-CD3 antibody. Cells were incubated with 1 µg ml⁻¹ soluble human anti-CD28 antibody for 48 h with varying doses of QM385 as indicated in Extended Data Fig. 7c. Cells were then collected for liquid chromatography mass spectrometry (LC-MS) BH4 measurements. Similar experiments were performed on anti-CD3/28-activated mouse splenocytes.

Plasma levels of QM385. To formulate 10 mg kg⁻¹ of QM385 for oral administration, we added 6.312 ml of 1% Tween-80 plus 0.5% hydroxypropyl methylcellulose (HPMC) in 50 mM carbonate buffer (pH 9.0) into a tube containing 7.98 mg QM385, then vortexed the mixture for 1–2 min and sonicated it for 30–35 min. Solutions were prepared just before use. The intravenous dosing solution was prepared in 10% dimethylacetamide (DMAC), 10% solutol HS15, 80% (10% (2-hydroxypropyl)-β-cyclodextrin) in saline (w/v). Approximately 100 µl

of blood sample was collected via the tail vein into EDTA-2K tubes. The blood samples were maintained in wet ice first, and centrifuged to obtain plasma (2,000g, 4 °C, 5 min) within 15 min of sampling. An aliquot of 30 µl serum sample was added with 100 µl acetonitrile containing 10 ng ml⁻¹ dexamethasone. The mixture was vortexed for 2 min and centrifuged at 14,000 r.p.m. for 5 min. An aliquot of 10 µl supernatant was injected for LC-MS/MS analysis. We also prepared a calibration curve of 0.100–1,000 ng ml⁻¹ for QM385 in diluted blood from C57BL/6 mice.

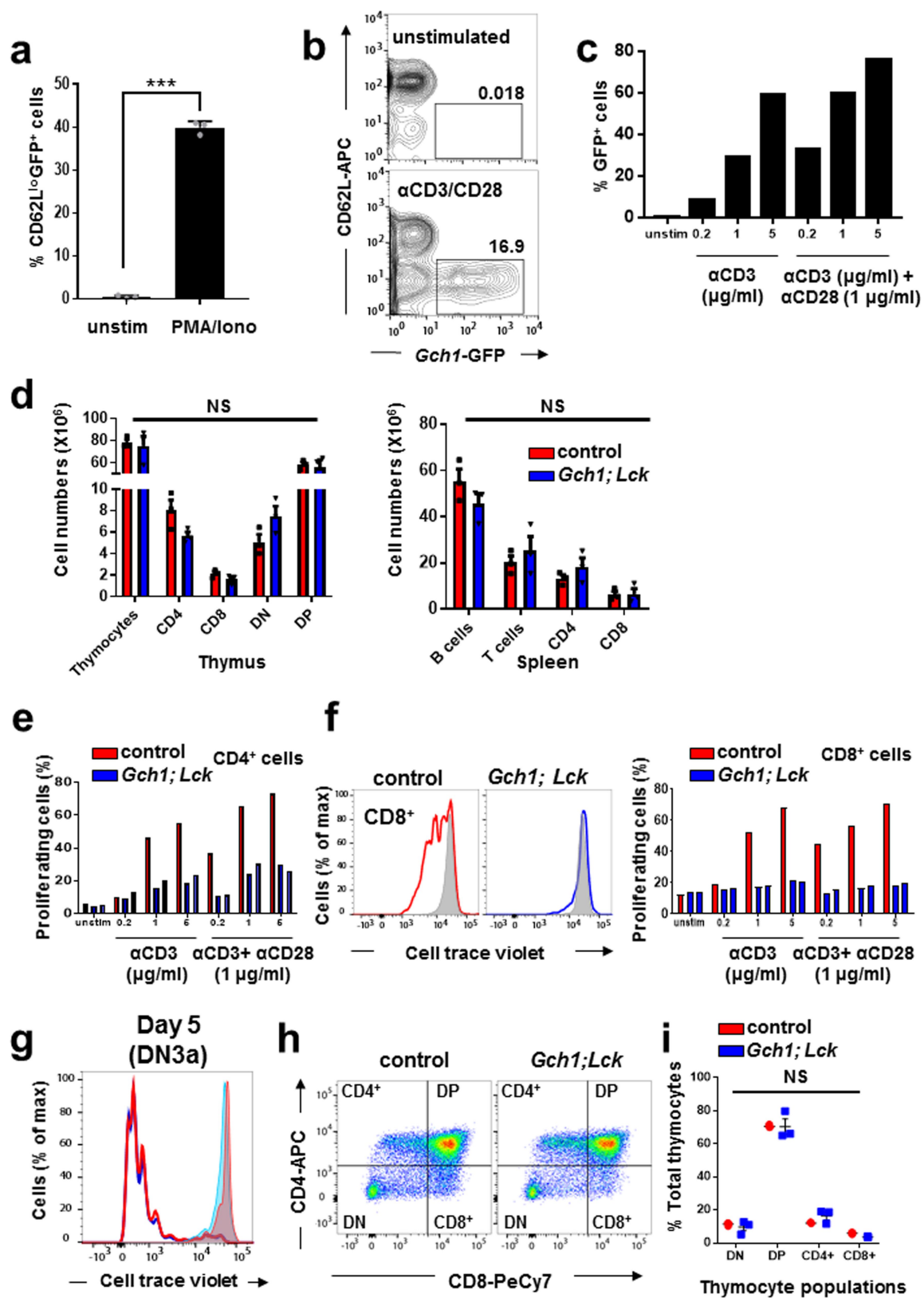
Statistical analyses. All values are expressed as means ± s.e.m. Details of the statistical tests used are stated in the figure legends. In brief, Student's *t*-test was used to compare between two groups. One-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons was used for analysis between multiple groups. Two-way ANOVA was used to compare two groups over time. In all tests, $P \leq 0.05$ was considered significant.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The microarray dataset is accessible through GEO accession number GSE108101. All other datasets generated and/or analysed during this study are available from the corresponding authors upon reasonable request.

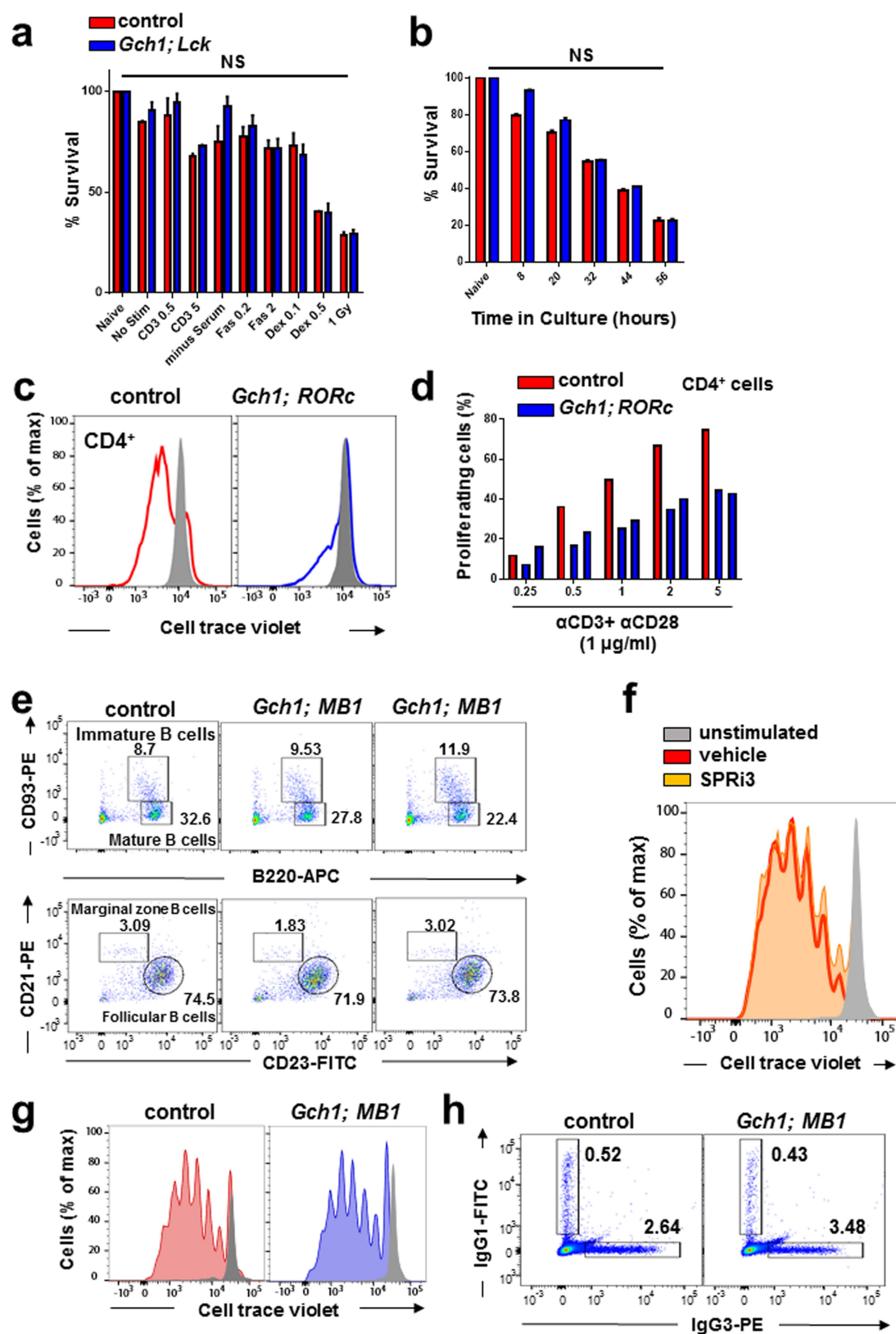
28. Hennen, T., Hagen, F. K., Tabak, L. A. & Marth, J. D. T-cell-specific deletion of a polypeptide N-acetylgalactosaminyl-transferase gene by site-directed recombination. *Proc. Natl Acad. Sci. USA* **92**, 12070–12074 (1995).
29. Sawada, S., Scarborough, J. D., Killeen, N. & Littman, D. R. A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* **77**, 917–929 (1994).
30. Ventura, A. et al. Restoration of p53 function leads to tumour regression *in vivo*. *Nature* **445**, 661–665 (2007).
31. Crabtree, M. J. et al. Quantitative regulation of intracellular endothelial nitric-oxide synthase (eNOS) coupling by both tetrahydrobiopterin-eNOS stoichiometry and biopterin redox status: insights from cells with tet-regulated GTP cyclohydrolase I expression. *J. Biol. Chem.* **284**, 1136–1144 (2009).
32. Banerjee, A. et al. Cellular and site-specific mitochondrial characterization of vital human amniotic membrane. *Cell Transplant.* **27**, 3–11 (2018).
33. Schmitt, T. M. & Zúñiga-Pflücker, J. C. Induction of T cell development from hematopoietic progenitor cells by delta-like-1 *in vitro*. *Immunity* **17**, 749–756 (2002).
34. Becker, C. et al. In vivo imaging of colitis and colon cancer development in mice using high resolution chromoendoscopy. *Gut* **54**, 950–954 (2005).
35. Collison, L. W. & Vignali, D. A. A. In vitro Treg suppression assays. *Methods Mol. Biol.* **707**, 21–37 (2011).
36. Boivin, N., Baillargeon, J., Doss, P. M. I. A., Roy, A. P. & Rangachari, M. Interferon-β suppresses murine Th1 cell function in the absence of antigen-presenting cells. *PLoS ONE* **10**, e0124802 (2015).
37. Lin, K. Y. et al. Treatment of established tumors with a novel vaccine that enhances major histocompatibility class II presentation of tumor antigen. *Cancer Res.* **56**, 21–26 (1996).
38. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, <https://doi.org/10.2202/1544-6115.1027> (2004).
39. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
40. Arai, N., Narisawa, K., Hayakawa, H. & Tada, K. Hyperphenylalaninemia due to dihydropteridine reductase deficiency: diagnosis by enzyme assays on dried blood spots. *Pediatrics* **70**, 426–430 (1982).
41. Theurl, I. et al. On-demand erythrocyte disposal and iron recycling requires transient macrophages in the liver. *Nat. Med.* **22**, 945–951 (2016).



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Upregulation of *Gch1* and BH4 in activated T cells. **a**, Percentage of CD62L^{lo} GFP⁺ cells from purified *Gch1-Gfp* CD4⁺ T cells stimulated for 24 h with phorbol myristate acetate and ionomycin (50 ng ml⁻¹ each). Data are shown as means \pm s.e.m., from $n = 3$ samples. The experiment was repeated two independent times. *** $P < 0.001$ (two-tailed Student's *t*-test). **b**, **c**, Representative *Gch1-Gfp* expression in 16-h-activated (CD62L^{low}) CD4⁺ T cells after anti-CD3/CD28 stimulation (**b**) and representative dose-response of anti-(α)CD3/CD28 stimulation of purified CD4⁺ *Gch1-Gfp* T cells for 24 h (**c**). The experiment was repeated two independent times with similar results. **d**, Cell numbers of various immune populations in the thymus (left) and spleen (right) from control ($n = 3$) and *Gch1;Lck* ($n = 3$) 8-week-old mice. Data from individual mice

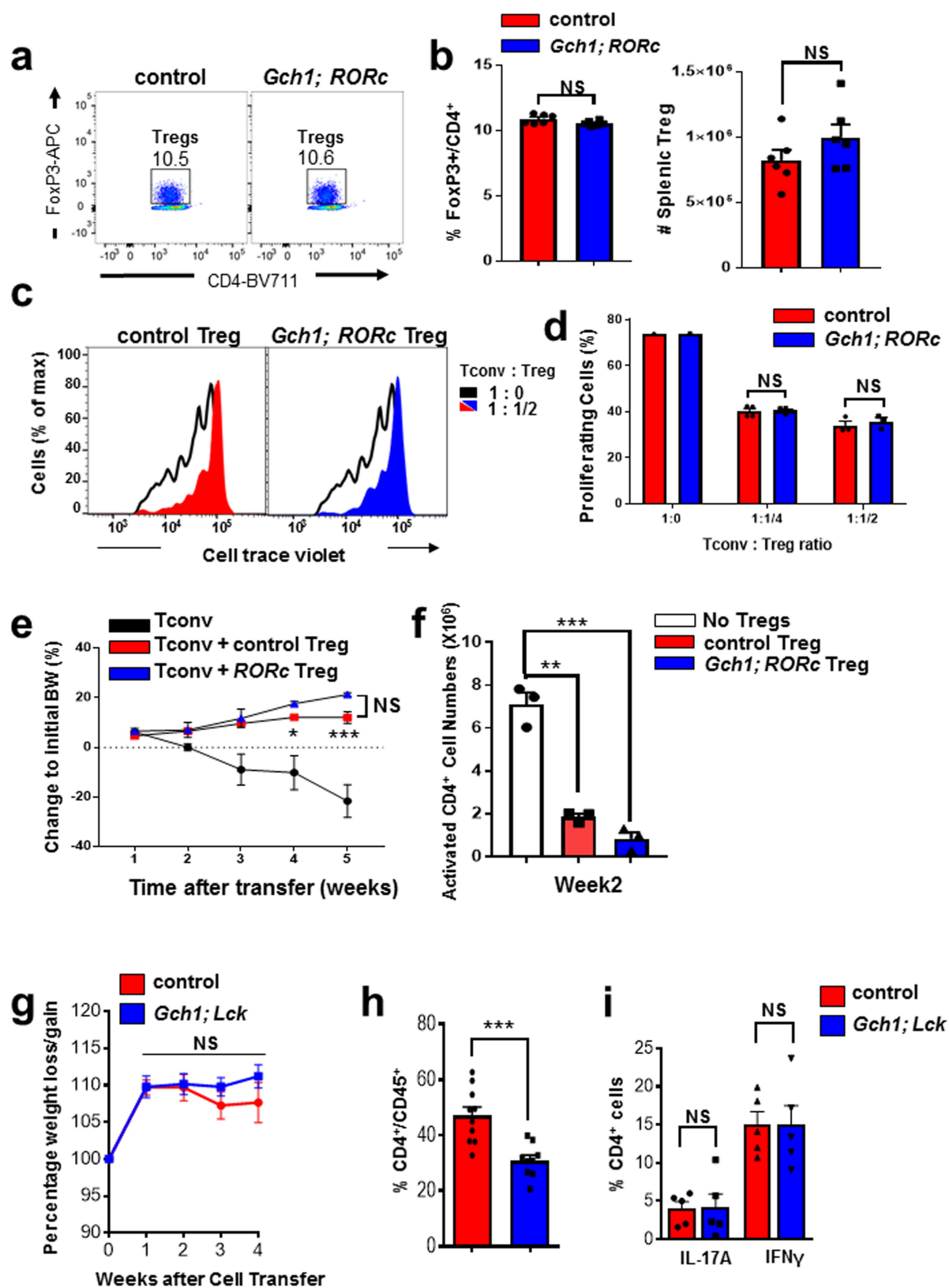
are shown as means \pm s.e.m. NS, not significant (two-tailed Student's *t*-test). **e**, **f**, CD4⁺ (**e**) and CD8⁺ (**f**) T cell proliferation after three days of anti-CD3/28 stimulation, from control and *Gch1;Lck* mice. **g**, Representative histogram depicting the proliferation of DN3a thymocytes from control and *Gch1;Lck* mice cultured on OP9-D11 stromal cells for five days. The experiment was repeated two independent times with similar results. **h**, **i**, Representative FACS blot depicting the differentiation into CD4⁺ and CD8⁺ T cells of DN3a thymocytes from control and *Gch1;Lck* mice cultured on OP9-D11 stromal cells for five days (**h**), and quantification of the differentiated cell types from $n = 3$ animals (**i**). Data from individual mice are shown as means \pm s.e.m. NS, not significant (two-tailed Student's *t*-test).



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Normal T cell development and B cell biology in the absence of *Gch1*. **a**, Thymocyte cell death induced over 24 h by various stimuli: anti-CD3 ($0.5 \mu\text{g ml}^{-1}$ and $5 \mu\text{g ml}^{-1}$), Fas ligand ($0.2 \mu\text{g ml}^{-1}$ and $2 \mu\text{g ml}^{-1}$), dexamethasone (Dex, $0.1 \mu\text{g ml}^{-1}$ and $0.5 \mu\text{g ml}^{-1}$) and γ -irradiation (1 Gray (Gy)). Data are shown as means \pm s.e.m. $n = 3$ for each genotype. NS, not significant (two-tailed Student's *t*-test). **b**, Death by neglect of purified CD4^+ T cells cultured without stimulation for up to 56 h. Data are shown as means \pm s.e.m. $n = 3$ for each genotype. NS, not significant (two-tailed Student's *t*-test). **c**, **d**, Proliferation of CD4^+ T cells from control and *Gch1*;ROR α mice after three days of anti-CD3/28 stimulation. Panels show representative FACS proliferation traces (**c**) and representative dose response (**d**). Experiments were repeated independently more than six times with similar results. **e**, Representative

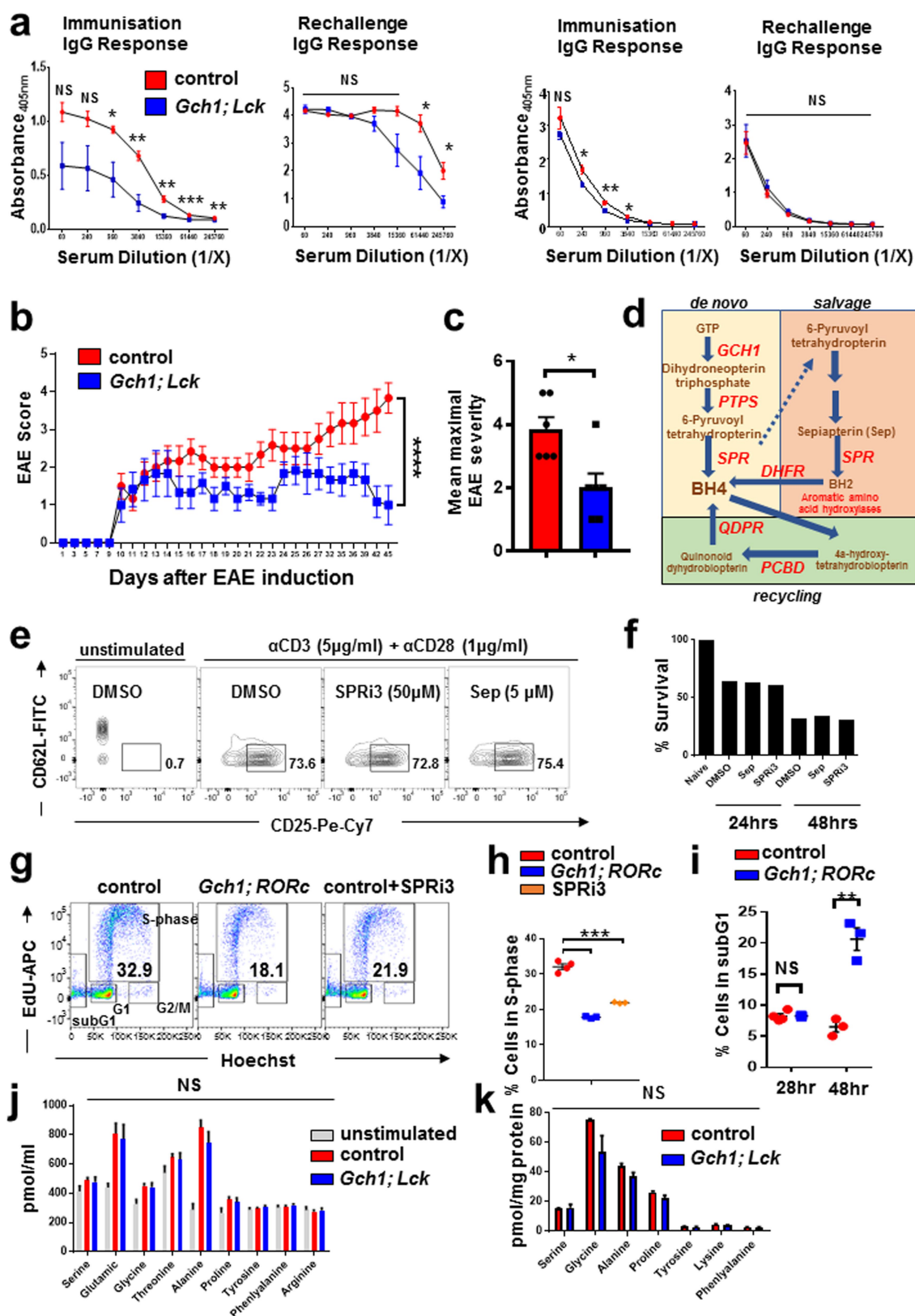
FACS plots from spleens of control and *Gch1*;MB1 mice. MB1-Cre is an early B cell deleter Cre line using endogenous *CD79a* B cell specific expression. The experiment was repeated two independent times with similar results. **f**, **g**, Representative FACS histogram depicting the proliferation of wild-type B cells treated with vehicle (DMSO) or SPRi3 ($50 \mu\text{M}$) (**f**), and of B cells from control and *Gch1*;MB1 mice in response to LPS ($1 \mu\text{g ml}^{-1}$) after three days (**g**). Shaded grey peaks represent unstimulated cells. FACS plots are representative of two independent experiments showing similar results. $n = 3$ mice per group. **h**, Class-switch recombination. FACS analysis of splenic CD43^+ B cells from control and *Gch1*;MB1 mice stimulated with LPS ($20 \mu\text{g ml}^{-1}$) for five days to induce class-switch recombination to IgG3. FACS plots are representative of two independent experiments showing similar results.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Development of regulatory T cells and their function in *Gch1*-ablated mice. **a, b,** Representative FACS plot depicting CD4⁺FoxP3⁺ regulatory T cells (T regs; **a**) and quantification of T-reg proportions as well as absolute numbers in the spleen (**b**) of control and *Gch1;RORc* mice ($n = 6$ each). Data are shown as means \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant (two-tailed Student's t -test). **c, d,** In vitro T-reg suppression assay, in which naive, wild-type CD4⁺ T cells were activated in the presence of varying ratios of T-reg cells from control and *Gch1;RORc* mice for four days. Representative histogram showing the suppressive capacity of control and *Gch1;RORc* T-reg cells (**c**) and quantification of proliferation with various ratios of T-reg cells (**d**). $n = 4$ samples. Data are shown as means \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; NS, not significant (two-tailed Student's t -test with multiple comparisons). Tconv, conventional CD4⁺ T cells (CD4⁺, CD25⁻ CD45RB^{high}). **e,** Naive CD4⁺ transfer colitis model, with co-transfer of FACS-purified T-reg cells from control ($n = 4$) and *Gch1;RORc* ($n = 4$) mice. As a control, Tconv cells (from $n = 16$ mice) with no co-transfer of T-reg cells were used. Changes

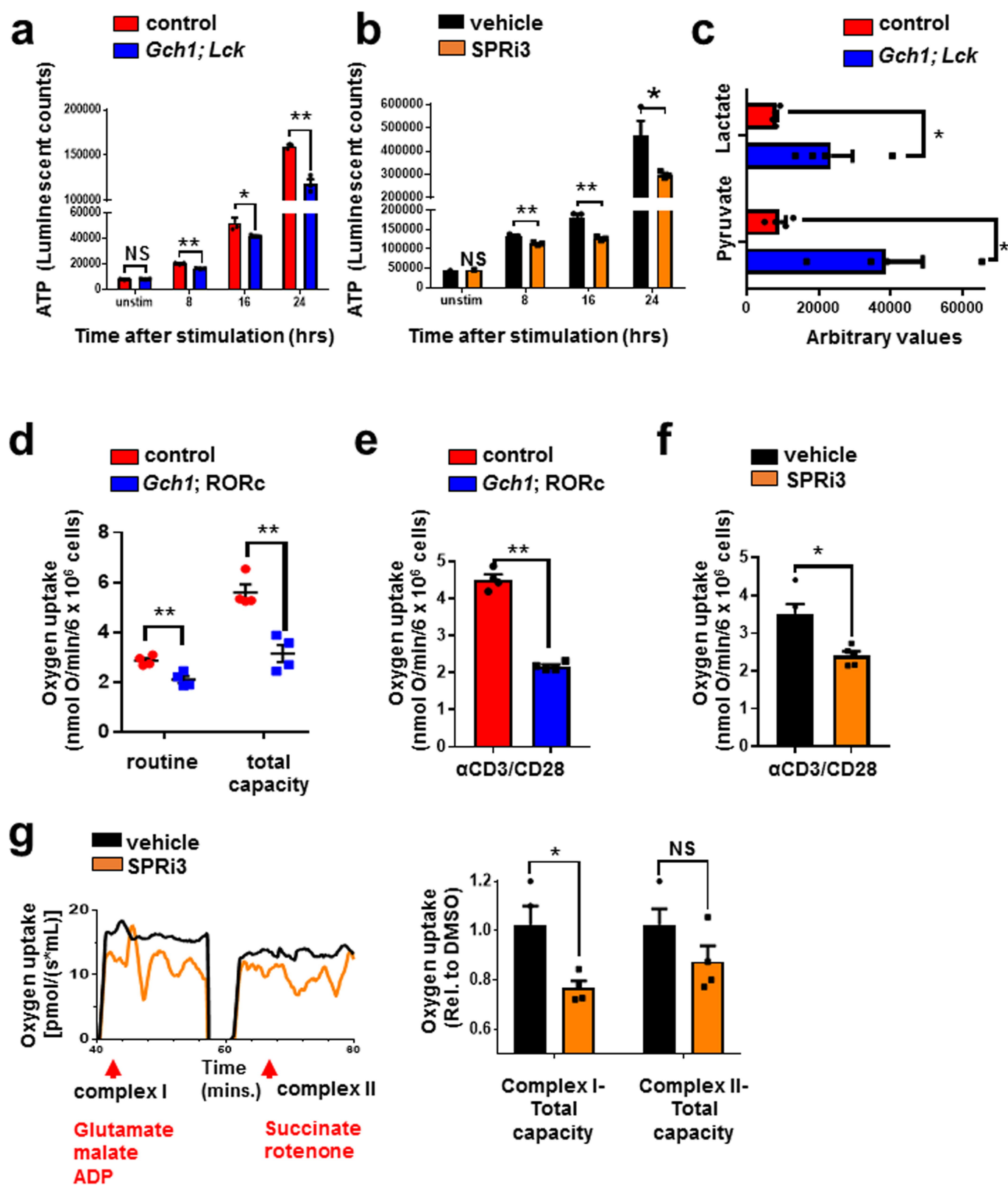
to initial body weight (BW) were scored over five weeks. Data are shown as means \pm s.e.m. * $P < 0.05$; *** $P < 0.001$; NS, not significant (two-way ANOVA with Tukey's multiple comparison test). **f,** Total numbers of CD4⁺ splenic T cells at two weeks post-transfer in mice ($n = 3$) transferred with naive CD4⁺ cells only ('no T regs') and mice transferred with T regs from control or *Gch1*-ablated (*Gch1;RORc*) mice. Data are shown as means \pm s.e.m. *** $P < 0.001$; **** $P < 0.0001$; NS, not significant (one-way ANOVA with Dunnett's multiple comparison test). **g,** Transfer colitis model of intestinal autoimmunity. Body-weight changes are plotted relative to initial weight in mice transferred with naive CD4⁺ T cells from control or *Gch1;Lck* mice ($n = 10$ each). Data are shown as means \pm s.e.m. NS, not significant (two-way ANOVA with Sidak's multiple comparisons). **h, i,** Proportion of CD4⁺ T cells in the draining mesenteric lymph nodes in week 4 (**h**), and profiles of intracellular cytokines (IFN- γ and IL-17) from transferred control and *Gch1;Lck* cells (**i**). Data are shown as means \pm s.e.m. $n = 10$ for each genotype for **h** and $n = 5$ for each genotype for **i**. *** $P < 0.001$; NS, not significant (two-tailed Student's t -test).



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Blockage of GCH1–BH4 abrogates T-cell-mediated autoimmunity. **a**, OVA immunization of control and *Gch1*;*Lck* mice. T-cell-dependent IgG responses and T-cell-independent IgM responses are shown two weeks after OVA immunization (left panels, 100 µg OVA in 200 µg alum) as well as two weeks after re-challenge (right panels). $n = 5$ for control mice; $n = 6$ for *Gch1*;*Lck* mice. Data are shown as means \pm s.e.m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; NS, not significant (two-tailed Student's *t*-test with multiple comparisons). **b**, **c**, EAE model of autoimmunity towards the central nervous system. Data are shown as means \pm s.e.m. **b**, EAE scores of control and *Gch1*;*Lck* mice. $n = 6$ for each genotype. $****P < 0.0001$ (linear regression analysis was performed on the slope of each curve). **c**, Mean maximal EAE severity in control and littermate *Gch1*;*Lck* mice. $*P < 0.05$ (Mann–Whitney test). **d**, Schematic of the de novo, salvage and recycling arms of the BH4 pathway. The dotted arrow indicates non-enzymatic reactions; solid arrows indicate enzymatic reactions. DHFR, dihydrofolate reductase; GTP, guanosine triphosphate; PCDB, pterin-4 α -carbinolamine dehydratase; PTPS, 6-pyruvoyl tetrahydropterin synthase; QDPR, quinoid dihydropteridine reductase; SPR, sepiapterin reductase. **e**, Representative FACS plots depicting activation marker profiles of purified wild-type control CD4⁺ T cells left

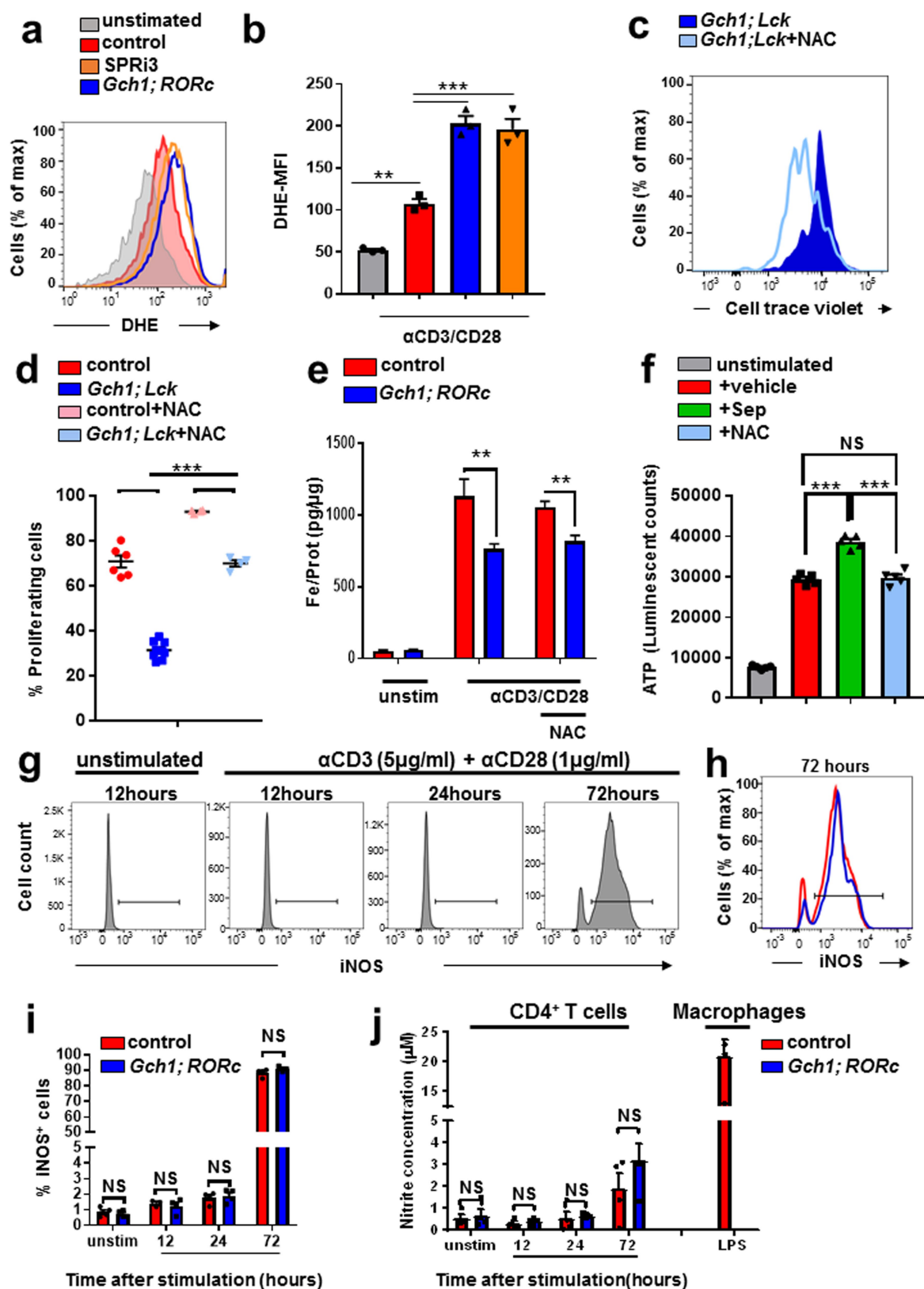
unstimulated or stimulated with anti-CD3/28 antibodies for 16 h and then treated with vehicle (DMSO), SPRI3 (50 µM) or sepiapterin (5 µM). The experiment was repeated two independent times with similar results. **f**, Cell survival as defined by the percentage of DAPI⁺annexinV⁺ cells from purified CD4⁺ T cells stimulated for 24 h or 48 h with anti-CD3/28 antibodies and then treated with vehicle (DMSO), SPRI3 (50 µM) or sepiapterin (5 µM). The experiment was repeated two independent times with similar results. **g**, **h**, Representative FACS blots depicting EdU cell-cycle analysis after 28 hours anti-CD3/CD28 stimulation of control, *Gch1*;*RORc* and SPRI3-treated control CD4⁺ T cells. EdU was pulsed for the last 4 hours (**g**) and quantification of S-phase entry (**h**). Data from individual mice are shown \pm s.e.m. $***P < 0.001$ (one-way ANOVA with Dunnett's multiple comparisons test). **i**, Quantification of subG1 (dead cells) populations after 24- and 48-h stimulation. EdU was pulsed for the last 4 hours of each time point. Data from individual mice are shown \pm s.e.m.). $**P < 0.01$; NS, not significant (multiple *t*-test comparisons). **j**, **k**, Amino acid profiles in the supernatants (**j**) and cell pellets (**k**) from 24-h anti-CD3/CD28-stimulated CD4⁺ T cells from control and *Gch1*;*Lck* mice. $n = 3$ for each genotype. Data are shown as means \pm s.e.m. NS, not significant (two-tailed Student's *t*-test).



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Mitochondrial dysfunction in BH4-depleted T cells after activation. **a, b**, ATP measurements in control ($n=3$) and *Gch1;Lck* ($n=3$) CD4⁺ T cells (**a**) and in wild-type CD4⁺ T cells treated with DMSO vehicle ($n=3$) or SPRI3 (50 μ M; $n=3$) (**b**), either left unstimulated or assayed at the indicated time points after T cell activation with anti-CD3/28 antibodies. Data are shown as means \pm s.e.m. $n=3$ for each genotype. * $P<0.05$; ** $P<0.01$ (two-tailed Student's t -test with multiple comparisons). **c**, Metabolomic measurements of lactate and pyruvate levels in cell pellets of 16-h anti-CD3/28-activated CD4⁺ T cells from control and *Gch1;Lck* mice. Data are shown as means \pm s.e.m. $n=4$ for each genotype. * $P<0.05$ (two-tailed Student's t -test). **d**, Routine and total capacitance oxygen respiration in intact, 16-h anti-CD3/CD28-stimulated CD4⁺ T cells from control and *Gch1;Lck* mice. Data from

individual mice are indicated \pm s.e.m. $n=4$ for each genotype. ** $P<0.01$ (two-tailed Student's t -test). **e, f**, Oxygen uptake rate in permeabilized, 16-h anti-CD3/CD28-stimulated CD4⁺ T cells from control ($n=4$) and *Gch1;RORc* ($n=4$) mice (**e**) and wild-type CD4⁺ T cells treated with DMSO or SPRI3 (50 μ M ($n=5$ each) (**f**). Data from individual mice are indicated \pm s.e.m. * $P<0.05$; ** $P<0.01$ (two-tailed Student's t -test). **g**, Left, representative oxygen consumption traces of complex-I-linked and complex-II-linked ETC activity from 16-h-activated wild-type CD4⁺ T cells treated with vehicle or SPRI3 (50 μ M). Right, relative complex-I- and complex-II-linked activities in activated control cells treated with vehicle ($n=4$) or SPRI3 (50 μ M; $n=4$). Data are shown as means \pm s.e.m. NS, not significant; * $P<0.05$ (two-tailed Student's t -test).

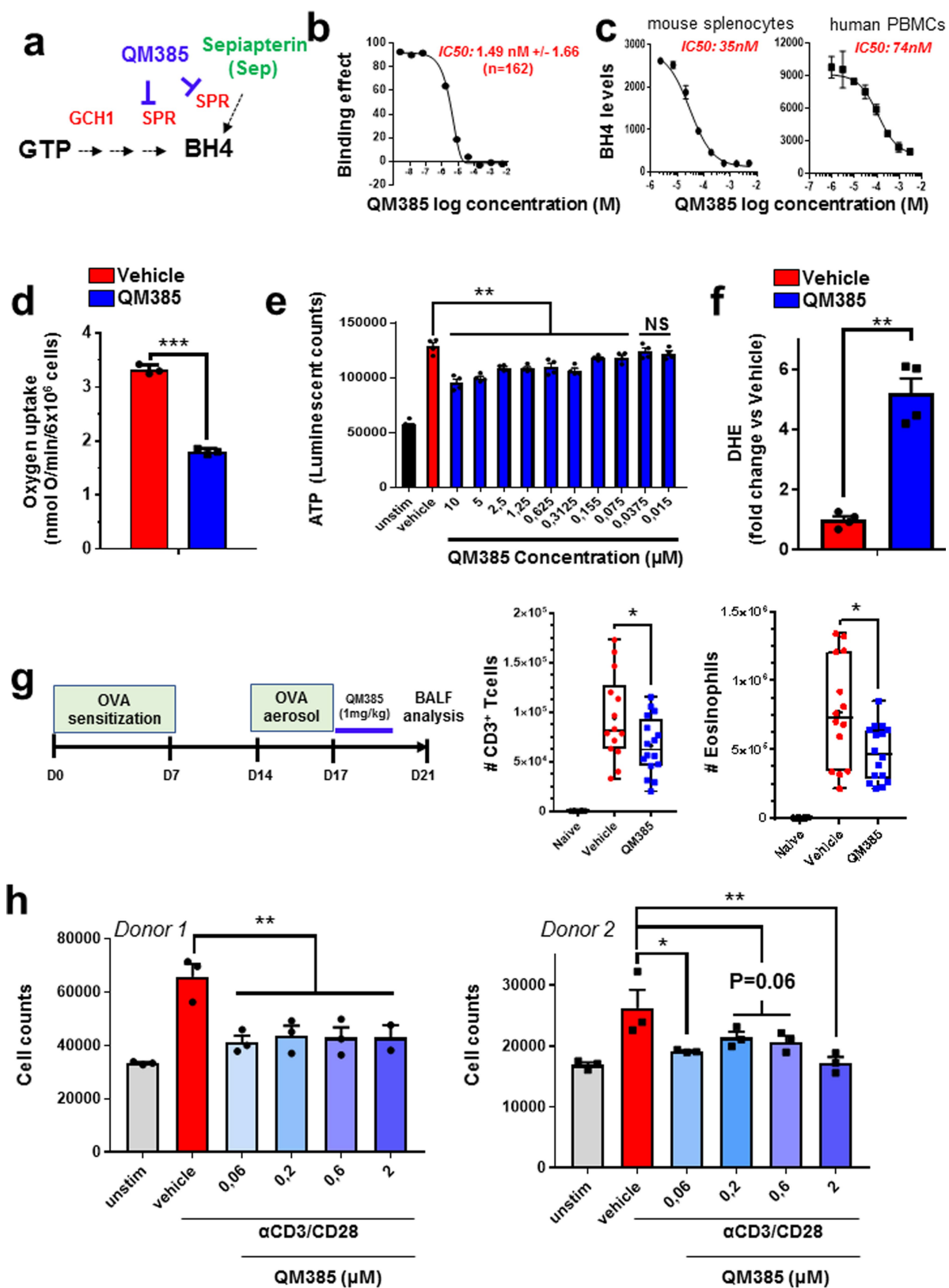


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Enhanced superoxide levels independent of iNOS coupling observed in BH4-deficient activated T cells.

a, b, Representative FACS histogram (**a**) and quantification of the mean fluorescent intensity (MFI; **b**) showing levels of DHE (dihydroethidium, a superoxide ROS indicator) in unstimulated and 20-h anti-CD3/28-activated CD4⁺ T cells from control and *GCH1;RORc* mice as well as control cells treated with SPRI3 (50 μ M). $n = 3$ samples per group. The experiment was repeated three independent times with similar results. **c, d**, Proliferation of control ($n = 6$) and *Gch1;Lck* ($n = 9$) CD4⁺ T cells and treatment with the superoxide scavenger NAC (500 μ M; $n = 4$ each). Representative three-day proliferation histograms are shown in **c**; quantification is shown in **d**. Data are given as means \pm s.e.m. Individual mice for each genotype are shown. **** $P < 0.0001$ (one-way ANOVA with Tukey's multiple comparison test). **e**, Total iron content from unstimulated or 24-h anti-CD3/28-stimulated CD4⁺ T cells (untreated or treated with 500 μ M NAC) from control ($n = 17, 4$, respectively) and *Gch1;RORc* ($n = 22, 6$, respectively) mice. Data are shown as means \pm s.e.m. Individual

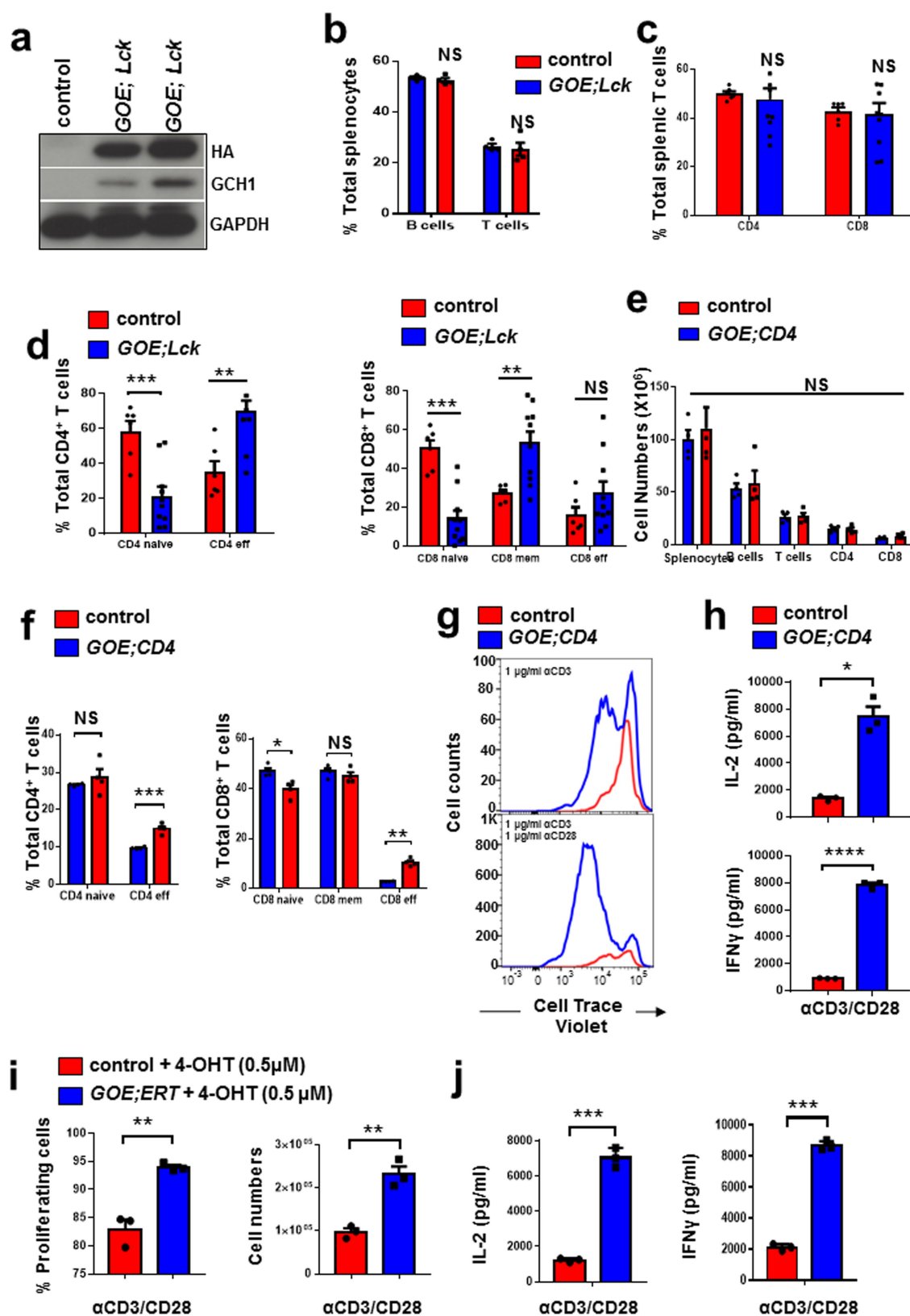
mice for each genotype are shown. ** $P < 0.01$ (two-tailed Student's *t*-test with Tukey's multiple comparisons). **f**, ATP measurements from stimulated wild-type CD4⁺ T cells treated with DMSO, sepiapterin and NAC for 24 h. Data are shown as means \pm s.e.m. $n = 5$ for each genotype. * $P < 0.05$; ** $P < 0.01$ (two-tailed Student's *t*-test with multiple comparisons). **g**, Intracellular iNOS expression in purified CD4⁺ control T cells left untreated or anti-CD3/CD28-stimulated for 12 h, 24 h or 72 h. The experiment was repeated two independent times with similar results. **h, i**, Representative histogram showing iNOS expression in control and *Gch1*-ablated CD4⁺ T cells stimulated with anti-CD3/CD28 antibodies for 72 h (**h**) and the percentage of iNOS⁺ cells was quantified over time (**i**). $n = 4$ for each genotype. Data are shown as means \pm s.e.m. NS, not significant (two-tailed Student's *t*-test). **j**, Nitrite measurements in the supernatant of stimulated cells from **i**. Peritoneal, thioglycollate-elicited macrophages stimulated with LPS (100 ng ml⁻¹) for 24 h were used as a positive control. Data are shown as means \pm s.e.m. $n = 4$ for each genotype. NS, not significant (two-tailed Student's *t*-test).



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Functional evaluation of the SPR blocker QM385. **a**, The BH4 pathway, indicating how QM385 acts on SPR, limiting BH4 production and correspondingly increasing sepiapterin levels, which can be used as a biomarker for QM385-mediated SPR inhibition. **b, c**, A representative concentration–response curve showing the binding affinity of QM385 to human SPR, tested in vitro by TR-FRET (**b**); and reduction of BH4 levels upon QM385 treatment in anti-CD3/28-stimulated mouse splenocytes (left panel, two independent experiments) and human PBMCs (right panel, two independent experiments) (**c**). The calculated half maximal inhibitory concentration (IC_{50}) values for each assay are indicated in red. The binding-effect assay was repeated 162 independent times with similar results. **d**, The oxygen-uptake rate in permeabilized, 16-h anti-CD3/CD28-stimulated wild-type $CD4^+$ T cells treated with DMSO or QM385 (2.5 μ M). Data from individual mice ($n = 3$) are indicated \pm s.e.m. $***P < 0.001$ (two-tailed Student's *t*-test). **e**, ATP measurements of unstimulated ($n = 8$) and 24-h-activated wild-type $CD4^+$ T cells treated with DMSO vehicle ($n = 4$) or varying doses

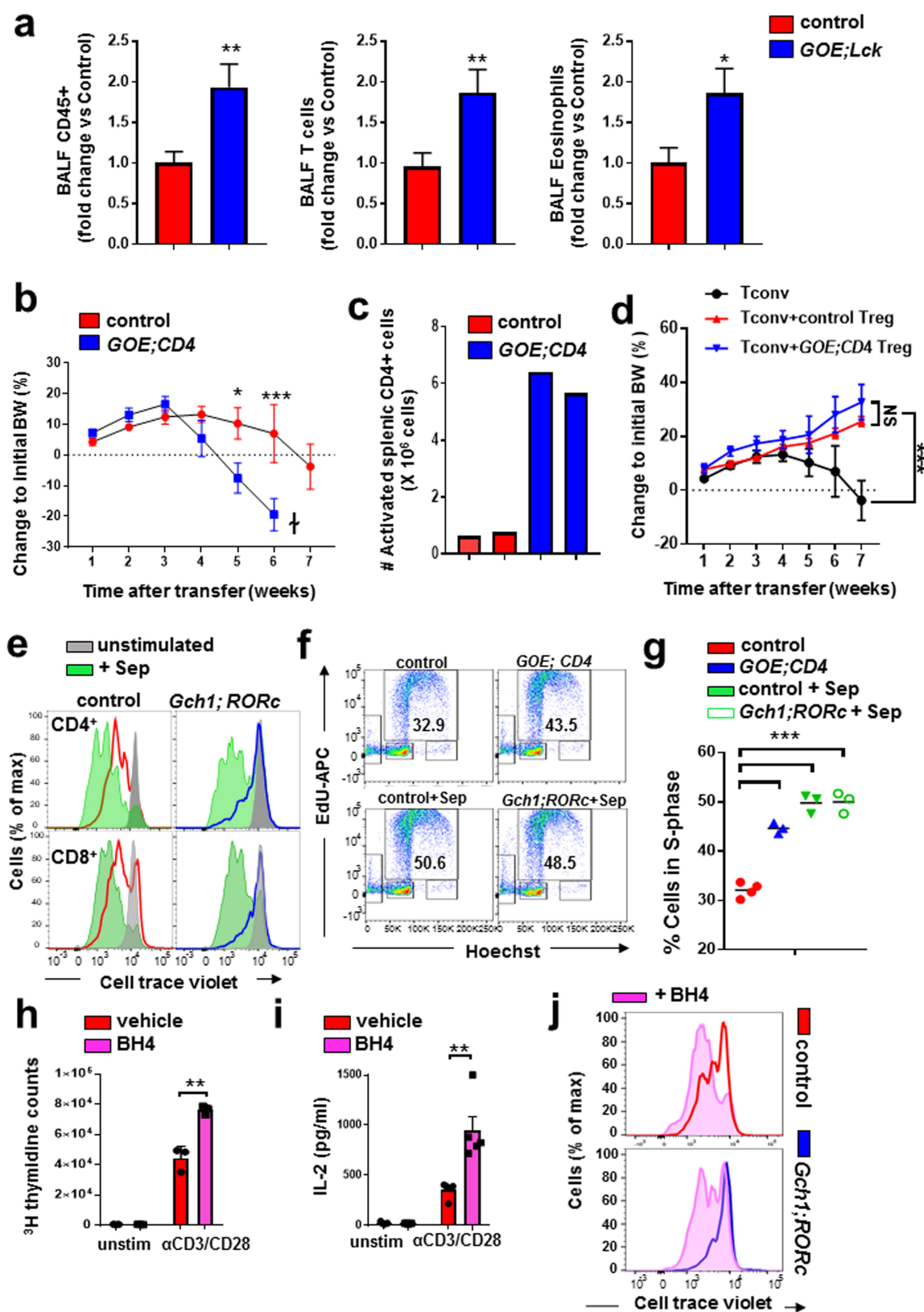
of QM385 ($n = 4$ for each dose). Data are shown as means \pm s.e.m. NS, not significant; $**P < 0.01$ (one-way ANOVA with Dunnett's multiple comparisons). **f**, Fold changes in DHE levels between $CD4^+$ T cells treated with DMSO or QM385 (2.5 μ M) and activated for 20 h. Data from individual mice ($n = 4$) are indicated \pm s.e.m. $**P < 0.01$ (two-tailed Student's *t*-test). **g**, Allergic airway inflammatory disease model and quantification of inflammatory cells in bronchoalveolar lavage fluids (BALFs). Data are shown as box-and-whisker plots (running from minimal to maximal values); individual data points are shown. $n = 15$ for vehicle-treated mice; $n = 17$ for QM385-treated mice. QM385 (1 mg kg^{-1}) was administered orally (peritoneally) twice a day for three consecutive days as depicted in the diagram. $*P < 0.05$; $**P < 0.01$ (two-tailed Student's *t*-test). **h**, Proliferation of human $CD4^+$ T cells from two donors performed in triplicate samples. Anti-CD3/28 T cells were stimulated with varying doses of QM385 and total counts were measured. Data are shown as means \pm s.e.m. $**P < 0.01$; $P < 0.05$ (one-way ANOVA with Dunnett's multiple comparisons).



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Increased numbers of effector T cells in naive mice overexpressing *Gch1*, and enhanced T cell proliferation after stimulation. **a**, Representative immunoblot to detect GCH1 and the HA tag in naive CD4⁺ T cells from control and *GOE;Lck*-overexpressing mice. The experiment was repeated three times with similar findings. **b, c**, The proportion of splenic T and B cells (**b**) and the proportion of CD4⁺ and CD8⁺ T cells among the splenic T cell (TCRβ⁺) population (**c**), from control ($n = 4$) and *GOE;Lck* ($n = 4$) mice. Data for individual mice aged eight weeks are shown as means \pm s.e.m. NS, not significant (two-tailed Student's *t*-test). **d**, Quantification of naive (CD44^{low}CD62L^{high}), memory (CD44^{high}CD62L^{high}) and effector (CD44^{high}CD62L^{low}) T cell subtypes from the spleen of control ($n = 6$) and *GOE;Lck* ($n = 10$) mice. Data for individual mice are shown as means \pm s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$; NS, not significant (two-tailed Student's *t*-test). **e**, Cell numbers for B cells, T cells, and CD4⁺ and CD8⁺ T cells in the spleens of control and *GOE;Cd4* mice. Data from individual mice ($n = 4$ for each genotype) are shown as means \pm s.e.m. NS, not significant (two-tailed Student's

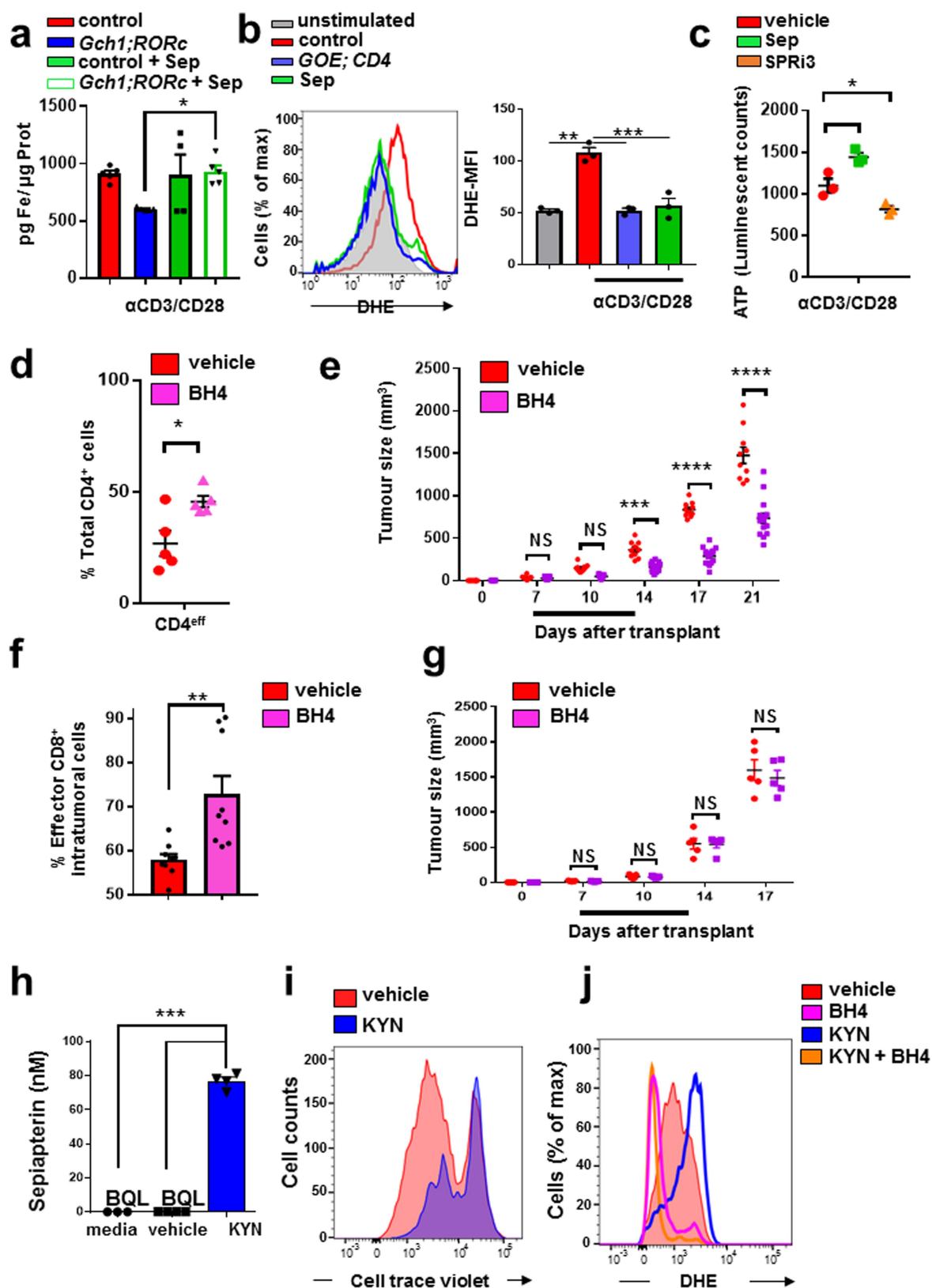
t-test). **f**, Proportion of CD4⁺ and CD8⁺ naive, memory and effector T cells in the spleens of naive control and *GOE;Cd4* mice. Data for individual mice ($n = 4$ for each genotype) are shown as means \pm s.e.m. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; NS, not significant (two-tailed Student's *t*-test). **g**, Representative histograms depicting dose-dependent proliferation of control and *GOE;Cd4* CD4⁺ T cells stimulated for three days with anti-CD3/28 antibodies. Experiments were repeated more than three times with comparable results. **h**, IL-2 and IFN- γ secretion after three days of stimulation (with anti-CD3/28 antibodies) of control and *GOE;Cd4* CD4⁺ T cells. Data are shown as means \pm s.e.m. $n = 3$ for each genotype. $^{*}P < 0.05$; $^{***}P < 0.0001$ (two-tailed Student's *t*-test). **i, j**, Cells from control ($n = 3$) and *GOE;ERT* ($n = 3$) mice were stimulated with anti-CD3/28 antibodies for three days and treated with 4-hydroxytamoxifen (4-OHT; 0.5 μ M) to induce *Gch1* overexpression in vitro. **i**, Quantification of proliferation of CD4⁺ T cells; **j**, cytokine secretion. Data from individual mice are shown as means \pm s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$ (two-tailed Student's *t*-test).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | T cells overproducing BH4 display enhanced ATP production, proliferation and autoimmunity. **a**, Allergic airway inflammatory disease model and fold change of inflammatory cells in BALFs, comparing control and *GOE;Lck* mice. Data are shown as means \pm s.e.m. $n = 18$ for control mice; $n = 17$ for *GOE;Lck* mice. $*P < 0.05$; $**P < 0.01$ (two-tailed Student's *t*-test). **b**, Transfer colitis model. Changes in body weight of *Rag2*^{-/-} mice transferred with control ($n = 6$ animals) or *GOE;Cd4* ($n = 5$) naive CD4⁺ T cells. Data are shown as means \pm s.e.m. $*P < 0.05$; $***P < 0.001$; NS, not significant (two-way ANOVA with Tukey's multiple comparison test). **c**, Total numbers of activated (CD62L^{low}CD44^{high}) CD4⁺ splenic T cells at three weeks post-transfer in mice transferred with control or *GOE;Cd4* naive CD4⁺ T cells. Data for two mice from each group are shown. **d**, Transfer colitis model, involving transfer of naive CD4⁺ T cells (150,000 cells) and co-transfer of FACS-purified T-reg cells from control ($n = 5$) and *GOE;Cd4* ($n = 6$) mice. Changes to initial body weights were scored over seven weeks. Data are shown as means \pm s.e.m. $***P < 0.001$; NS, not significant (two-way ANOVA with Tukey's multiple comparison test). **e**, Representative histograms depicting the proliferation of purified unstimulated and anti-CD3/28-stimulated CD4⁺ and CD8⁺ wild-type

and *Gch1;RORc* T cells treated for three days with sepiapterin (5 μ M). The profile for the unstimulated T cells of each genotype is shown in grey. Experiments were repeated three independent times with comparable results. **f**, **g**, Representative FACS plots showing EdU-based cell-cycle analysis following 28-h anti-CD3/28 stimulation of control CD4⁺ T cells, *GOE;Cd4* CD4⁺ T cells, control CD4⁺ T cells treated with sepiapterin (5 μ M), and *GCH1;RORc* CD4⁺ T cells treated with sepiapterin (5 μ M) (**f**); and quantification of the S-phase-entry population (**g**). EdU was pulsed for the last 4 h of stimulation. $n = 4$ mice for control; $n = 3$ mice for all other genotypes. $***P < 0.001$ (one-way ANOVA with Dunnett's multiple comparisons test). **h**, **i**, Effect of BH4 on the proliferation (³H-thymidine incorporation; **h**) and IL-2 secretion (**i**) of CD4⁺ wild-type T cells activated with anti-CD3/28 antibodies for 24 h and treated with vehicle ($n = 3/4$) or BH4 (10 μ M; $n = 3/4$). Data are shown for individual mice as means \pm s.e.m. $**P < 0.01$ (two-tailed Student's *t*-test). **j**, Representative histograms depicting the proliferation of control and *Gch1;RORc* CD4⁺ T cells after three days of anti-CD3/28 stimulation supplemented with BH4 (10 μ M). FACS blots are representative of two independent experiments with comparable results.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Overactivation of the GCH1–BH4 pathway leads to enhanced anti-tumour immunity. **a**, Total iron content from 24-h anti-CD3/28 stimulated CD4⁺ T cells (untreated or treated with 5 μ M sepiapterin) from control ($n = 5/4$) and *Gch1*;*RORc* ($n = 5$) mice. Data are shown as means \pm s.e.m.; individual mice for each genotype are shown. * $P < 0.05$ (one-way ANOVA with Tukey's multiple comparisons). **b**, Representative FACS histogram depicting DHE levels (left) and quantification of the mean fluorescent intensity (right) in unstimulated and 20-h anti-CD3/28-activated CD4⁺ T cells from control and *GOE*;*Cd4* littermates as well as wild-type cells treated with sepiapterin (5 μ M). $n = 3$ for each condition. Data are shown as means \pm s.e.m. ** $P < 0.01$; *** $P < 0.001$ (one-way ANOVA with Tukey's multiple comparisons test). **c**, ATP measurements for stimulated wild-type CD4⁺ T cells treated with DMSO, sepiapterin (5 μ M) or SPRI3 (50 μ M) for 24 h. Data are shown as means \pm s.e.m. $n = 3$ for each genotype. * $P < 0.05$; ** $P < 0.01$ (two-tailed Student's *t*-test with multiple comparisons). **d**, Quantification of intratumoral effector CD4⁺ T cells (CD44⁺CD62L^{low}) assayed from E0071 tumours on day 28 for vehicle- and BH4-treated mice. Data are shown as means \pm s.e.m. $n = 5$ mice for each condition. * $P < 0.05$; ** $P < 0.01$ (two-tailed Student's *t*-test). **e**, Effect of BH4 supplementation on H-Ras-transformed TC-1 tumour growth. TC-1 tumour cells were orthotopically injected; once the tumours were palpable (day 7), BH4 (35 mg kg⁻¹; $n = 15$) or vehicle (saline; $n = 10$) was therapeutically administered for seven days as indicated. Data are shown for individual mice as

means \pm s.e.m. *** $P < 0.001$; **** $P < 0.0001$ (two-way ANOVA with Sidak's multiple comparisons). **f**, Quantification of intratumoral effector CD8⁺ T cells (CD44⁺CD62L^{low}) assayed from TC-1 tumours on day 21 in vehicle- or BH4-treated mice ($n = 9$ mice for each genotype). Data are shown as means \pm s.e.m. ** $P < 0.01$ (two-tailed Student's *t*-test). **g**, Effect of BH4 supplementation on TC-1 tumour growth in *Rag2*^{-/-} hosts. TC-1 tumour cells were orthotopically injected into *Rag2*^{-/-} female mice; once the tumours were palpable (day 7), BH4 (35 mg kg⁻¹; $n = 5$) or vehicle (saline; $n = 5$) was administered. BH4 and vehicle supplementation was carried out for seven days as indicated on the graph. Data are shown for individual mice as means \pm s.e.m. NS, not significant (two-way ANOVA with Sidak's multiple comparisons). **h**, Sepiapterin levels in the supernatant of wild-type CD4⁺ T cells stimulated with anti-CD3/28 antibodies for 20 h and treated with vehicle or kynurenine (KYN; 150 μ M). Culture medium was also included for comparison. BQL, below quantifiable levels. Data are shown as means \pm s.e.m. $n = 4$ independent samples for each condition. *** $P < 0.001$ (one-way ANOVA with Tukey's multiple comparisons test). **i**, Representative histogram depicting proliferation of three-day anti-CD3/28-activated wild-type CD4⁺ T cells treated with vehicle or kynurenine (50 μ M). **j**, Representative FACS histograms depicting DHE levels in anti-CD3/28-stimulated wild-type CD4⁺ T cells treated with vehicle (DMSO), kynurenine alone (50 μ M) or kynurenine (50 μ M) plus BH4 (10 μ M) for 20 h. The experiment was repeated three independent times with comparable results.

Autophagy maintains tumour growth through circulating arginine

Laura Poillet-Perez¹, Xiaoli Xie¹, Le Zhan¹, Yang Yang¹, Daniel W. Sharp¹, Zhixian Sherrie Hu¹, Xiaoyang Su^{1,2}, Anurag Maganti¹, Cherry Jiang¹, Wenyun Lu³, Haiyan Zheng⁴, Marcus W. Bosenberg⁵, Janice M. Mehnert^{1,6}, Jessie Yanxiang Guo^{1,2,7}, Edmund Lattime^{1,8}, Joshua D. Rabinowitz^{1,3} & Eileen White^{1,9*}

Autophagy captures intracellular components and delivers them to lysosomes, where they are degraded and recycled to sustain metabolism and to enable survival during starvation^{1–5}. Acute, whole-body deletion of the essential autophagy gene *Atg7* in adult mice causes a systemic metabolic defect that manifests as starvation intolerance and gradual loss of white adipose tissue, liver glycogen and muscle mass¹. Cancer cells also benefit from autophagy. Deletion of essential autophagy genes impairs the metabolism, proliferation, survival and malignancy of spontaneous tumours in models of autochthonous cancer^{6,7}. Acute, systemic deletion of *Atg7* or acute, systemic expression of a dominant-negative ATG4b in mice induces greater regression of KRAS-driven cancers than does tumour-specific autophagy deletion, which suggests that host autophagy promotes tumour growth^{1,8}. Here we show that host-specific deletion of *Atg7* impairs the growth of multiple allografted tumours, although not all tumour lines were sensitive to host autophagy status. Loss of autophagy in the host was associated with a reduction in circulating arginine, and the sensitive tumour cell lines were arginine auxotrophs owing to the lack of expression of the enzyme argininosuccinate synthase 1. Serum proteomic analysis identified the arginine-degrading enzyme arginase I (ARG1) in the circulation of *Atg7*-deficient hosts, and in vivo arginine metabolic tracing demonstrated that serum arginine was degraded to ornithine. ARG1 is predominantly expressed in the liver and can be released from hepatocytes into the circulation. Liver-specific deletion of *Atg7* produced circulating ARG1, and reduced both serum arginine and tumour growth. Deletion of *Atg5* in the host similarly released circulating arginine and suppressed tumorigenesis, which demonstrates that this phenotype is specific to autophagy function rather than to deletion of *Atg7*. Dietary supplementation of *Atg7*-deficient hosts with arginine partially restored levels of circulating arginine and tumour growth. Thus, defective autophagy in the host leads to the release of ARG1 from the liver and the degradation of circulating arginine, which is essential for tumour growth; this identifies a metabolic vulnerability of cancer.

To validate whether host autophagy promotes tumour growth, we tested the growth of an autophagy-competent C57Bl/6J isogenic *Braf*^{V600E/+}*Pten*^{-/-}*Cdkn2a*^{-/-} mouse melanoma cell line (termed YUMM 1.1) in C57Bl/6J host mice, without (*Atg7*^{+/+}) and with (*Atg7*^{Δ/Δ}) conditional whole-body *Atg7* deficiency (Fig. 1a). YUMM 1.1 tumours were significantly smaller when grown in *Atg7*^{Δ/Δ} hosts compared to *Atg7*^{+/+} hosts (Fig. 1b), demonstrating that host autophagy promoted tumour growth. The examination of additional cell lines—autophagy-competent isogenic C57Bl/6J *Braf*^{V600E/+}*Pten*^{-/-}*Cdkn2a*^{-/-} YUMM 1.3 melanoma, carcinogen-induced MB49 urothelial

carcinoma and *Kras*^{G12D/+}*p53*^{-/-} (*p53* is also known as *Trp53*) 71.8 non-small-cell lung cancer cells—revealed a similar requirement for host autophagy for tumour growth (Extended Data Fig. 1a, c, e). The decreased tumour growth observed in *Atg7*^{Δ/Δ} hosts was associated with decreased proliferation. In some tumour types, there was also increased apoptosis (Fig. 1c, Extended Data Fig. 1b, d, f). However, host autophagy was not required for the growth of autophagy-competent isogenic C57Bl/6J *Braf*^{V600E/+}*Pten*^{-/-}*Cdkn2a*^{-/-} YUMM 1.7 and 1.9 melanoma cell lines (Extended Data Fig. 2a–d), which indicates that—although dependency on host autophagy is common—there are tumour-specific adaptation mechanisms.

The melanoma cell lines that are dependent on host autophagy for tumour growth are derived from genetically engineered mouse models of cancer, and therefore have a low mutation burden, low neoantigen load and fail to provoke an efficient T cell response⁹. Nonetheless, autophagy modulates a variety of immune mechanisms that could underlie defective tumour growth in autophagy-deficient hosts. *Atg7*^{Δ/Δ} hosts did not modify infiltration of YUMM 1.1 tumours with CD3⁺, CD4⁺ or CD8⁺ cells (Extended Data Fig. 2e). Depletion of CD4⁺ and CD8⁺ T cells modestly increased tumour growth in *Atg7*^{+/+} hosts but did not significantly rescue growth in *Atg7*^{Δ/Δ} hosts (Extended Data Fig. 2f). Thus, despite the relative increase in the fraction of myeloid-derived suppressor cells and CD8⁺ T cells in *Atg7*^{Δ/Δ} hosts (Extended Data Fig. 2g), the decreased tumour growth in *Atg7*^{Δ/Δ} hosts was not due to the induction of an anti-tumour T cell response.

Autophagy supports metabolism by recycling cargo to provide anabolic and catabolic substrates⁶. This metabolic recycling function of autophagy promotes mammalian survival during fasting^{1–3}, and tumour cell survival under conditions of nutrient limitation^{4,10,11}. One major source of tumour nutrients is the host blood supply. Accordingly, we tested whether circulating nutrients provided by host autophagy were required for tumour growth. Metabolite profiling of serum from *Atg7*^{+/+} and *Atg7*^{Δ/Δ} hosts identified 12 metabolites that were decreased and 7 that were increased with autophagy knockout (Fig. 1d, Supplementary Tables 1, 2). Serum arginine was notably downregulated in *Atg7*^{Δ/Δ} compared to *Atg7*^{+/+} hosts (−2.37 fold change (log₂(serum arginine in *Atg7*^{Δ/Δ}/serum arginine in *Atg7*^{+/+})) (Fig. 1d), confirming previous results¹.

Arginine is a non-essential amino acid derived from the diet, de novo synthesis and protein turnover, and is important for mTOR activation¹², ammonia detoxification through the urea cycle as well as the synthesis of proteins, creatine, polyamines and nitric oxide¹³. It has long been known that some human cancers silence expression of ASS1, the gene that encodes argininosuccinate synthase 1 (ASS1), which results in arginine auxotrophy¹⁴. Without ASS1, cancer cells are unable to synthesize arginine from citrulline and are dependent

¹Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA. ²Department of Medicine, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA. ³Department of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. ⁴Biological Mass Spectrometry Facility, Rutgers University, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA. ⁵Department of Pathology, Yale University School of Medicine, New Haven, CT, USA. ⁶Department of Medicine, Division of Medical Oncology, Developmental Therapeutics Unit, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA. ⁷Department of Chemical Biology, Rutgers Ernest Mario School of Pharmacy, Piscataway, NJ, USA. ⁸Department of Surgery, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ, USA. ⁹Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ, USA. *e-mail: epwhite@cinj.rutgers.edu

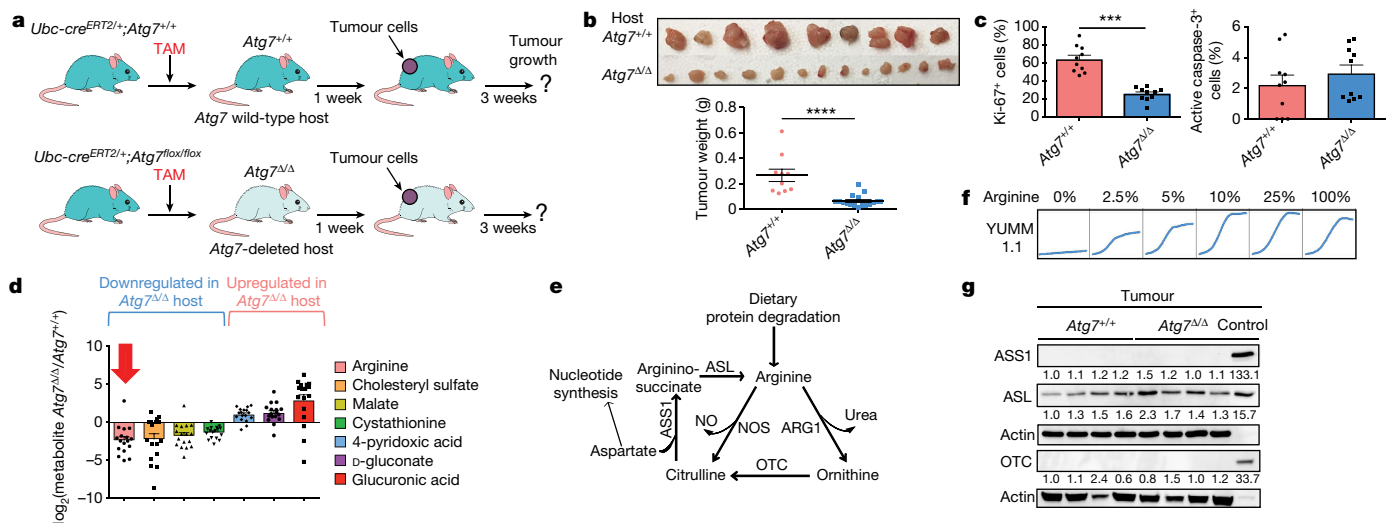


Fig. 1 | Host autophagy promotes growth of arginine auxotrophic tumours. **a**, Experimental design to induce host mice with conditional whole-body *Atg7* deletion (*Atg7^{Δ/Δ}*) and wild-type controls (*Atg7^{+/+}*) with which to assess tumour growth. *Ubc-cre^{ERT2/+};Atg7^{+/+}* and *Ubc-cre^{ERT2/+};Atg7^{Δ/Δ}* mice were injected with tamoxifen (TAM) to delete *Atg7* and were then injected subcutaneously with tumour cells. Tumour growth was monitored over three weeks. **b**, Comparison of tumour weight between *Atg7^{+/+}* ($n = 5$) and *Atg7^{Δ/Δ}* ($n = 8$) hosts. Data are mean \pm s.e.m. **** $P < 0.0001$. **c**, Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ cells in tumours from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts. Data are mean \pm s.e.m. *** $P < 0.001$. **d**, Serum metabolites with fold-change (\log_2 (metabolite in *Atg7^{Δ/Δ}*/metabolite in

Atg7^{+/+})) cut-offs of >1 or <-1 between *Atg7^{+/+}* ($n = 17$) and *Atg7^{Δ/Δ}* ($n = 17$) hosts obtained by liquid chromatography mass spectrometry (LC-MS), with $P < 0.05$. **e**, Illustration of the arginine metabolism. NO, nitric oxide; NOS, nitric oxide synthetase. **f**, YUMM 1.1 proliferation in vitro, in medium containing different percentages of arginine. Cell density was measured every 2 h using IncuCyte. Data are representative of three independent experiments performed in duplicate. **g**, Western blotting showing expression of ASS1, ASL and OTC in tumours from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts ($n = 4$ each), representative of three independent experiments. The kidney was used as a control tissue for ASS1 and ASL, and the liver was used for OTC. Actin was used as a loading control. In all figures, n represents the number of mice.

on exogenous arginine^{15,16}. ASS1 silencing prevents consumption of aspartate by the urea cycle, increasing the availability of this amino acid—which is required for pyrimidine biosynthesis and can become

limiting in hypoxia^{17,18} (Fig. 1e). These findings suggested that low circulating arginine may underlie defective tumour growth in autophagy-deficient hosts.

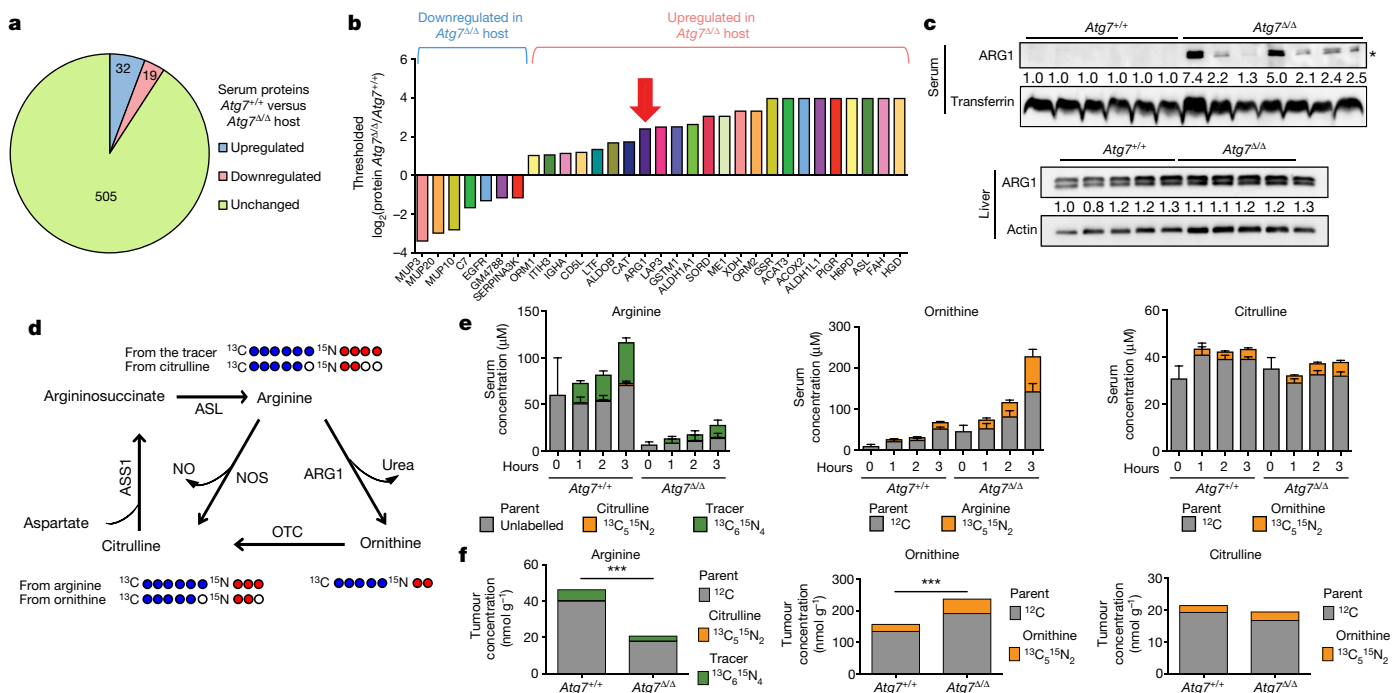


Fig. 2 | Levels of ARG1 in serum increase in *Atg7^{Δ/Δ}* hosts and deplete circulating arginine. **a**, Comparison of serum proteins between *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts ($n = 5$ each) obtained by nano LC-MS/MS with corrected $P < 0.05$. **b**, Proteins with fold-change (\log_2 (protein in *Atg7^{Δ/Δ}*/protein in *Atg7^{+/+}*)) cut-offs of >1 or <-1 between *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts. **c**, Western blotting showing expression of ARG1 in serum and liver from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts. * $P < 0.05$ compared to *Atg7^{+/+}* hosts. Data are representative of two independent experiments.

Actin and transferrin were used as loading controls. **d**, Illustration of the labelling pattern of the $^{13}\text{C}_6^{15}\text{N}_4$ arginine-tracer. **e**, Concentration (in μM) of arginine, citrulline and ornithine in serum from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts ($n = 3$ and 4, respectively) after infusion with $^{13}\text{C}_6^{15}\text{N}_4$ -arginine. Data are mean \pm s.e.m. **f**, Concentration (in nmol g^{-1}) of arginine, citrulline and ornithine in tumours from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts ($n = 2$ each) after infusion with $^{13}\text{C}_6^{15}\text{N}_4$ -arginine. Data are mean. *** $P < 0.001$ by two-way ANOVA test.

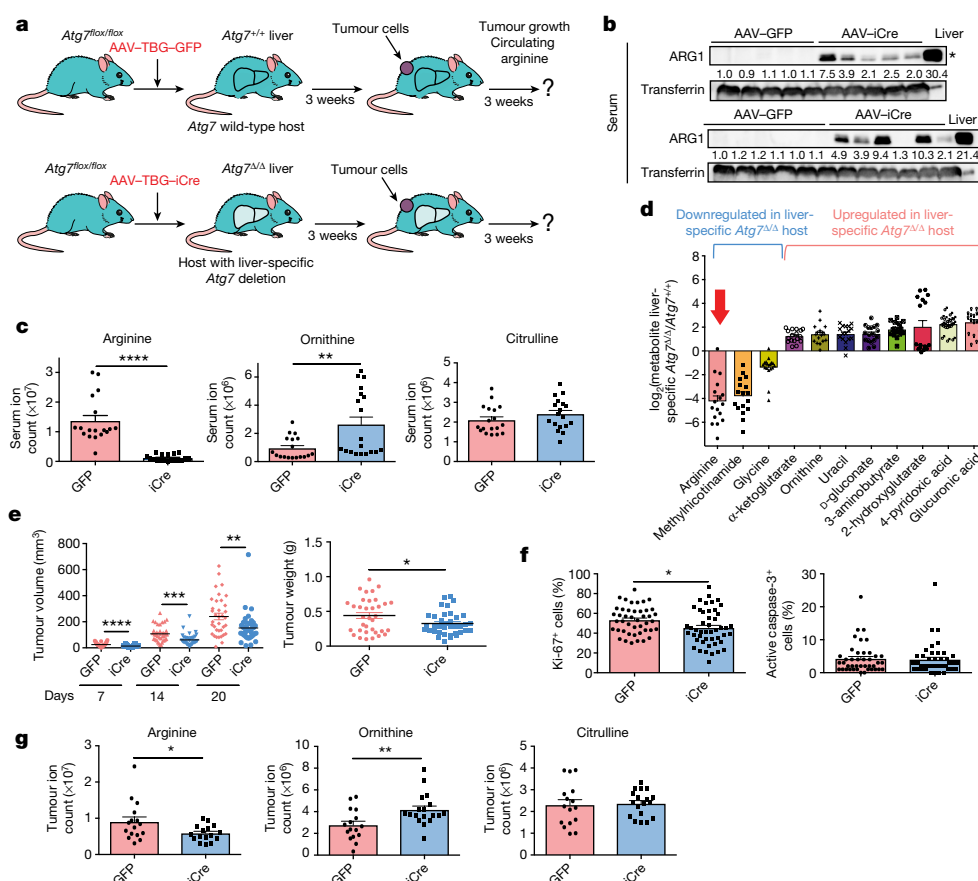


Fig. 3 | *Atg7* deletion in liver increases serum ARG1, and decreases serum arginine and tumour growth. **a**, Experimental design to induce liver-specific deletion of *Atg7*. *Atg7^{flax/flax}* mice were injected in the tail vein with AAV-TBG-GFP or AAV-TBG-iCre to delete *Atg7* in the liver, and were injected subcutaneously with tumour cells. Tumour growth was monitored over three weeks. **b**, Western blotting showing expression of ARG1 in serum from *Atg7^{+/+}* hosts and hosts with liver-specific deletion of *Atg7* ($n = 11$ each). * $P < 0.05$ compared to *Atg7^{+/+}* hosts. Data are representative of two independent experiments. Transferrin was used as a loading control. **c**, Levels of arginine, ornithine and citrulline in serum in *Atg7^{+/+}* hosts and hosts with liver-specific deletion of *Atg7* ($n = 18$ each), obtained by LC-MS. Data are mean \pm s.e.m. ** $P < 0.01$, **** $P < 0.0001$.

d, Serum metabolites with fold-change ($\log_2(\text{metabolite in liver-specific } Atg7^{\Delta/\Delta}/\text{metabolite in } Atg7^{+/+})$) cut-offs of >1 or <-1 between *Atg7^{+/+}* hosts and hosts with liver-specific deletion of *Atg7* ($n = 17$ each) obtained by LC-MS, with $P < 0.05$. Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **f**, Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ cells in tumours from *Atg7^{+/+}* and liver-specific *Atg7^{Δ/Δ}* hosts. Data are mean \pm s.e.m. * $P < 0.05$. **g**, Levels of arginine, ornithine and citrulline in tumours in *Atg7^{+/+}* hosts ($n = 16$) and hosts with liver-specific deletion of *Atg7* ($n = 16$), obtained by LC-MS. Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$.

To determine their requirement for exogenous arginine, YUMM 1.1, 1.3, 1.7, 1.9, MB49 and 71.8 cells were tested for growth without and with arginine. Proliferation was blocked in vitro in complete medium with the sole absence of arginine, and this was not associated with cell death. Growth rates increased with an increased percentage of arginine in the medium, which demonstrates arginine auxotrophy (Fig. 1f, Extended Data Fig. 3a). YUMM 1.1 tumours were tested for the lack of expression of enzymes involved in arginine biosynthesis: ASS1, argininosuccinate lyase (ASL)—which converts citrulline to arginine—and ornithine transcarbamylase (OTC), which converts ornithine to citrulline (Fig. 1e). As previously shown for melanoma^{19,20}, and irrespective of the use of *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts, tumours lacked ASS1 and OTC expression, which explains their arginine auxotrophy (Fig. 1g). In contrast to tumours, both *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts express ASS1, ASL and OTC in liver and ASS1 and ASL in kidney, which suggests that they are capable of arginine synthesis (Extended Data Fig. 3b). Consistent with findings from the YUMM 1.1 line and the literature^{14–16}, YUMM 1.7 tumours that grew on *Atg7^{Δ/Δ}* hosts also lacked expression of ASS1 and OTC, which suggests that—in a subset of tumour cell lines—there is a mechanism of intrinsic resistance that is independent of arginine auxotrophy (Extended Data Fig. 3c).

To determine how circulating arginine is depleted in *Atg7^{Δ/Δ}* hosts, we examined the serum proteome by nanoscale liquid chromatography

coupled to tandem mass spectrometry (nano LC-MS/MS), which identified 19 proteins that were downregulated and 32 that were upregulated upon loss of *Atg7* (Fig. 2a, Supplementary Table 3). ARG1 was among the proteins that were upregulated in the serum of *Atg7^{Δ/Δ}* hosts (2.43 fold change ($\log_2(\text{ARG1 in } Atg7^{\Delta/\Delta}/\text{ARG1 in } Atg7^{+/+})$) (Fig. 2b). ARG1 is expressed in the liver, where it degrades arginine to ornithine. The appearance of ARG1 in serum, without altered levels in the liver, in *Atg7^{Δ/Δ}* hosts was confirmed by western blotting (Fig. 2c). Levels of nitric oxide in serum were not modified, which suggests that serum ARG1 did not alter the arginine availability for nitric oxide synthesis (Extended Data Fig. 3d). Serum ARG1 activity in vitro was increased, as shown by greater rates of degradation of ¹³C₆-arginine to ¹³C₅-ornithine in serum from *Atg7^{Δ/Δ}* hosts (Extended Data Fig. 4a). To determine how *Atg7* deficiency altered arginine metabolism in vivo, we infused *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts with ¹³C₆¹⁵N₄-labelled arginine, for three hours²¹ (Fig. 2d). To analyse ¹³C and ¹⁵N enrichment, serum was collected at different times during infusion, and tumours, kidneys and livers were collected at the end of the three-hour infusion. The serum of *Atg7^{Δ/Δ}* hosts showed decreased arginine (¹²C, ¹³C₆¹⁵N₄ and ¹³C₅¹⁵N₂) associated with increased ornithine (¹²C and ¹³C₅¹⁵N₂), which indicates degradation of circulating arginine to ornithine (Fig. 2e, Extended Data Fig. 4b). Kidneys from *Atg7^{Δ/Δ}* hosts showed decreased levels of arginine, with no change in ornithine or citrulline; no difference was

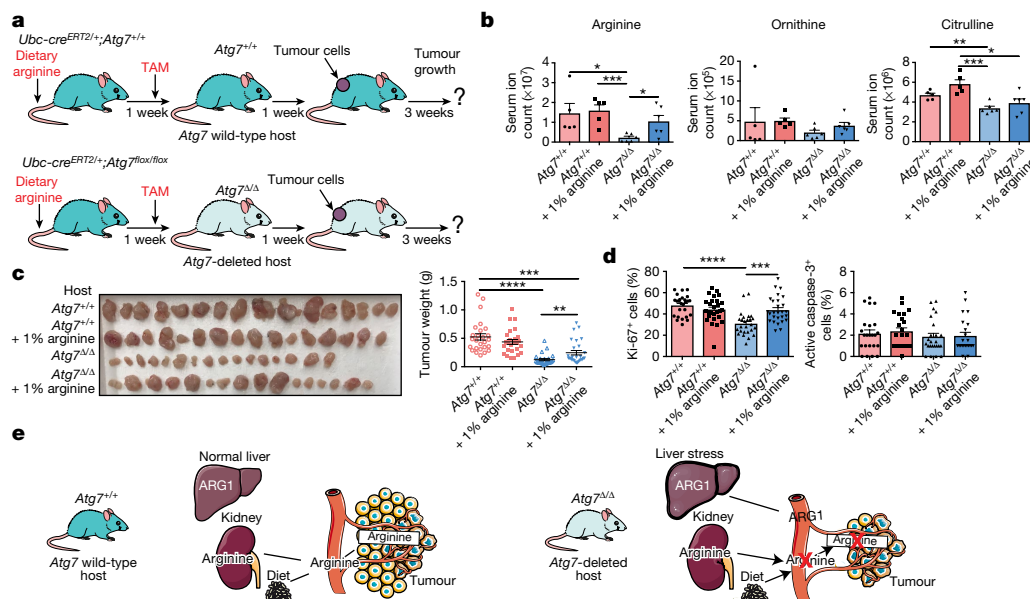


Fig. 4 | Dietary arginine supplementation rescues tumour growth in *Atg7* Δ/Δ hosts. **a**, Experimental design to perform arginine supplementation and induce conditional whole-body deletion of *Atg7* to assess YUMM 1.1 tumour growth. *Ubc-cre^{ERT2/+};Atg7^{+/+}* and *Ubc-cre^{ERT2/+};Atg7^{lox/flox}* mice were supplied with supplementary dietary arginine (0 or 1%). Seven days later, tamoxifen was injected to delete *Atg7* and mice were injected subcutaneously with tumour cells. Tumour growth was monitored over three weeks. **b**, Serum arginine, ornithine and citrulline in *Atg7^{+/+}* ($n = 5$) and *Atg7^{Δ/Δ}* ($n = 6$) hosts with or without arginine supplementation, obtained by LC–MS. Data are mean \pm s.e.m.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **c**, Comparison of tumour weight between *Atg7^{+/+}* ($n = 13$), *Atg7^{+/+}* + 1% arginine ($n = 13$), *Atg7^{Δ/Δ}* ($n = 13$) and *Atg7^{Δ/Δ}* + 1% arginine ($n = 14$) hosts. Data are mean \pm s.e.m. ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **d**, Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ cells in tumours from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts, with or without arginine supplementation. Data are mean \pm s.e.m. *** $P < 0.001$, **** $P < 0.0001$. **e**, Model of host autophagy promoting tumour growth. Illustrations are by Servier (https://smart.servier.com/), CC-BY-3.0.

observed in levels of arginine, citrulline or ornithine in the livers of *Atg7^{Δ/Δ}* hosts compared to *Atg7^{+/+}* hosts (Extended Data Fig. 4c, d). Tumours from *Atg7^{Δ/Δ}* hosts also displayed decreased levels of arginine (^{12}C , $^{13}\text{C}_6^{15}\text{N}_4$ and $^{13}\text{C}_5^{15}\text{N}_2$) and increased levels of ornithine (^{12}C and $^{13}\text{C}_5^{15}\text{N}_2$) (Fig. 2f). These results confirm that arginine is depleted in ASS1-deficient tumours that are dependent on exogenous arginine, which is consistent with insufficient circulating arginine in *Atg7^{Δ/Δ}* hosts.

During inflammation, injury and liver disease, ARG1 is released from hepatocytes into circulation, which leads to arginine depletion²². *Atg7^{Δ/Δ}* hosts have steatosis¹, and liver-specific deletion of *Atg5* or *Atg7* is associated with liver damage^{2,23,24}. Accordingly, we hypothesized that ARG1 is released into circulation after deletion of *Atg7* in the liver. To test this hypothesis, we deleted *Atg7* specifically in the liver and examined levels of arginine and ARG1 in circulation, and tumour growth (Fig. 3a). Injection of an AAV-TBG-iCre vector efficiently deleted *Atg7* in the liver, but not in other organs such as the brain and kidney (Extended Data Fig. 5a–c). As expected, liver-specific deletion of *Atg7* led to histopathologic changes in liver cells without affecting other tissues (Extended Data Fig. 5d). As seen in *Atg7^{Δ/Δ}* hosts, serum from hosts with liver-specific deletion of *Atg7* showed increased levels of ARG1 (Fig. 3b), a reduced level of arginine, an increased level of ornithine (Fig. 3c) and no change in the level of nitric oxide (Extended Data Fig. 5e). Liver-specific deletion of *Atg7* also modified levels of other circulating metabolites, with 18 increased and 4 decreased compared to *Atg7^{+/+}* hosts (Supplementary Tables 4, 5). Some of these circulating metabolites (for example, 4-pyridoxic acid, D-glucuronate and glucuronic acid) were also altered in *Atg7^{Δ/Δ}* hosts, which suggests that the dysregulation of these metabolites has a liver-specific origin (Fig. 1d, Extended Data Fig. 5f). The level of arginine in serum was downregulated in hosts with liver-specific deletion of *Atg7* compared to *Atg7^{+/+}* hosts (Fig. 3d), as was shown in *Atg7^{Δ/Δ}* hosts (Fig. 1d). The weight and volume of YUMM 1.1 melanoma tumours were significantly decreased in hosts with liver-specific deletion of *Atg7*, compared to *Atg7^{+/+}* hosts (Fig. 3e); this decrease in weight and volume

was associated with decreased proliferation and no change in apoptosis (Fig. 3f). Tumours from hosts with liver-specific deletion of *Atg7* had decreased levels of arginine and increased levels of ornithine, but to a lesser extent than the tumours from *Atg7^{Δ/Δ}* hosts—this may explain why the decreased tumour growth in hosts with liver-specific deletion of *Atg7* was not as marked as with *Atg7^{Δ/Δ}* hosts (Fig. 3g). In hosts with liver-specific deletion of *Atg7*, autophagy in the microenvironment may locally feed the tumour with amino acids, as has previously been shown in pancreatic cancer and *Drosophila* tumours^{25,26}. These results suggest that deletion of *Atg7* in the liver is responsible for ARG1 release into the circulation, which leads to depletion of circulating arginine and decreased tumour growth.

To determine whether the degradation of circulating arginine by ARG1 in *Atg7^{Δ/Δ}* hosts was due to loss of autophagy, we examined mice with conditional deletion of *Atg5*. Whole-body conditional deletion of *Atg5* also introduced ARG1 into circulation and decreased the level of arginine in serum; tumour growth was also decreased in these *Atg5^{Δ/Δ}* hosts (Extended Data Fig. 6). Similar to liver-specific deletion of *Atg7*, liver-specific deletion of *Atg5* led to histopathologic changes in the liver with increased circulating ARG1 and reduced arginine (Extended Data Fig. 7), which confirms that the modulation of circulating arginine and tumorigenesis was dependent on autophagy.

We next tested whether dietary arginine supplementation can rescue tumour growth in *Atg7^{Δ/Δ}* hosts (Fig. 4a). Dietary arginine supplementation was able to partially increase levels of arginine in serum in *Atg7^{Δ/Δ}* hosts, and did not modify levels of ornithine or citrulline (Fig. 4b). This increased circulating arginine promoted growth and proliferation of the YUMM 1.1 and 1.3 melanoma cell lines in *Atg7^{Δ/Δ}* hosts, compared to *Atg7^{+/+}* hosts (Fig. 4c, d, Extended Data Fig. 8a–c), confirming that limiting circulating arginine can curtail tumour growth.

In summary, autophagy in the liver prevents the release of ARG1 and the degradation of circulating arginine that is important for the growth of arginine-auxotrophic tumours (Fig. 4e). However, some tumour cells that are auxotrophic for arginine in vitro were capable of growth in *Atg7^{Δ/Δ}* hosts, suggesting that adaptation mechanisms exist²⁷. Recent

work has demonstrated that autophagy in the local tumour microenvironment can provide amino acids that promote tumour growth^{25,26}. Our work demonstrates that host autophagy also sustains a circulating amino acid—arginine—that is essential for tumour growth. This finding underscores the importance of understanding the sensitivity of ASS1-deficient tumours to arginine deprivation therapy²⁸, with or without autophagy inhibition²⁹. As tumour nutrients are mainly derived from host circulation, restricting essential tumour nutrients in the circulation—as done with asparaginase treatment—is a form of cancer therapy that is ripe for further exploitation³⁰.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0697-7>.

Received: 3 May 2018; Accepted: 17 September 2018;

Published online 14 November 2018.

- Karsli-Uzunbas, G. et al. Autophagy is required for glucose homeostasis and lung tumor maintenance. *Cancer Discov.* **4**, 914–927 (2014).
- Komatsu, M. et al. Impairment of starvation-induced and constitutive autophagy in *Atg7*-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
- Kuma, A. et al. The role of autophagy during the early neonatal starvation period. *Nature* **432**, 1032–1036 (2004).
- Guo, J. Y. et al. Autophagy provides metabolic substrates to maintain energy charge and nucleotide pools in Ras-driven lung cancer cells. *Genes Dev.* **30**, 1704–1717 (2016).
- Kamada, Y., Sekito, T. & Ohsumi, Y. Autophagy in yeast: a TOR-mediated response to nutrient starvation. *Curr. Top. Microbiol. Immunol.* **279**, 73–84 (2004).
- Kimmelman, A. C. & White, E. Autophagy and tumor metabolism. *Cell Metab.* **25**, 1037–1043 (2017).
- Amaravadi, R., Kimmelman, A. C. & White, E. Recent insights into the function of autophagy in cancer. *Genes Dev.* **30**, 1913–1930 (2016).
- Yang, A. et al. Autophagy sustains pancreatic cancer growth through both cell-autonomous and nonautonomous mechanisms. *Cancer Discov.* **8**, 276–287 (2018).
- Wang, J. et al. UV-induced somatic mutations elicit a functional T cell response in the YUMMER1.7 mouse melanoma model. *Pigment Cell Melanoma Res.* **30**, 428–435 (2017).
- Strohecker, A. M. et al. Autophagy sustains mitochondrial glutamine metabolism and growth of *Braf*^{V600E}-driven lung tumors. *Cancer Discov.* **3**, 1272–1285 (2013).
- Guo, J. Y. et al. Autophagy suppresses progression of K-ras-induced lung tumors to oncocytomas and maintains lipid homeostasis. *Genes Dev.* **27**, 1447–1461 (2013).
- Chantranupong, L. et al. The CASTOR proteins are arginine sensors for the mTORC1 Pathway. *Cell* **165**, 153–164 (2016).
- Morris, S. M. Jr. Arginine metabolism: boundaries of our knowledge. *J. Nutr.* **137**, 1602S–1609S (2007).
- Delage, B. et al. Arginine deprivation and argininosuccinate synthetase expression in the treatment of cancer. *Int. J. Cancer* **126**, 2762–2772 (2010).
- Dillon, B. J. et al. Incidence and distribution of argininosuccinate synthetase deficiency in human cancers: a method for identifying cancers sensitive to arginine deprivation. *Cancer* **100**, 826–833 (2004).
- Patil, M. D., Bhaumik, J., Babykutty, S., Banerjee, U. C. & Fukumura, D. Arginine dependence of tumor cells: targeting a chink in cancer's armor. *Oncogene* **35**, 4957–4972 (2016).
- Rabinovich, S. et al. Diversion of aspartate in ASS1-deficient tumours fosters de novo pyrimidine synthesis. *Nature* **527**, 379–383 (2015).
- Nagamani, S. C. & Erez, A. A metabolic link between the urea cycle and cancer cell proliferation. *Mol. Cell. Oncol.* **3**, e1127314 (2016).
- Feun, L. G. et al. Negative argininosuccinate synthetase expression in melanoma tumours may predict clinical benefit from arginine-depleting therapy with pegylated arginine deiminase. *Br. J. Cancer* **106**, 1481–1485 (2012).
- Lam, T. L. et al. Recombinant human arginase inhibits the in vitro and in vivo proliferation of human melanoma by inducing cell cycle arrest and apoptosis. *Pigment Cell Melanoma Res.* **24**, 366–376 (2011).
- Hui, S. et al. Glucose feeds the TCA cycle via circulating lactate. *Nature* **551**, 115–118 (2017).
- Morris, S. M. Jr. Arginases and arginine deficiency syndromes. *Curr. Opin. Clin. Nutr. Metab. Care* **15**, 64–70 (2012).
- Takamura, A. et al. Autophagy-deficient mice develop multiple liver tumors. *Genes Dev.* **25**, 795–800 (2011).
- Komatsu, M. et al. Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. *Cell* **131**, 1149–1163 (2007).
- Sousa, C. M. et al. Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion. *Nature* **536**, 479–483 (2016).
- Katheder, N. S. et al. Microenvironmental autophagy promotes tumour growth. *Nature* **541**, 417–420 (2017).
- Kremer, J. C. et al. Arginine deprivation inhibits the Warburg effect and upregulates glutamine anaplerosis and serine biosynthesis in ASS1-deficient cancers. *Cell Reports* **18**, 991–1004 (2017).
- Yau, T. et al. A phase 1 dose-escalating study of pegylated recombinant human arginase 1 (Peg-rhArg1) in patients with advanced hepatocellular carcinoma. *Invest. New Drugs* **31**, 99–107 (2013).
- Shen, W. et al. A novel and promising therapeutic approach for NSCLC: recombinant human arginase alone or combined with autophagy inhibitor. *Cell Death Dis.* **8**, e2720 (2017).
- Koprivnikar, J., McCloskey, J. & Faderl, S. Safety, efficacy, and clinical utility of asparaginase in the treatment of adult patients with acute lymphoblastic leukemia. *Onco Targets Ther.* **10**, 1413–1422 (2017).

Acknowledgements This work was supported by National Institutes of Health grants: R01CA130893, R01CA188096 (to E.W.), R01CA163591 (to E.W. and J.D.R.), K22CA190521 (to J.Y.G.), R50CA211437 (to W.L.), R01CA193970 and the V Foundation for Cancer Research (to J.M.M.). L.P.-P. received support from a postdoctoral fellowship from the New Jersey Commission for Cancer Research (DHFS16PPC034). We thank the Rutgers-New Brunswick/Robert Wood Johnson Medical School Biological Mass Spectrometry Facility (S100D016400) for mass spectrometry analysis, and the Biospecimen Repository and Histopathology Service, Metabolomics Service, Flow Cytometry and Biometrics Shared Resources (D. Moore performed the statistical analysis of the proteomics data) of Rutgers Cancer Institute of New Jersey (P30CA072720).

Reviewer information Nature thanks R. DeBerardinis and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions L.P.-P. performed the majority of the experimental work and wrote the manuscript. L.Z. performed surgery and infusion with labelled arginine. Y.Y. developed the methods and provided the mice required for generating *Atg5*^{Δ/Δ} and hosts with liver-specific deletion of *Atg5*. A.M. and C.J. assisted with in vitro experiments. X.X. and J.Y.G. performed some of the tumour growth experiments. J.M.M. provided melanoma expertise. D.W.S. and E.L. assisted with CD4 and CD8 depletion. Z.S.H. assisted with mouse husbandry. H.Z. performed proteomics processing and analysis. X.S., W.L. and J.D.R. performed metabolomics processing and analysis. M.W.B. provided YUMM 1.1, 1.3, 1.7 and 1.9 melanoma cells. E.W. is the leading principal investigator who conceived the project, supervised research and edited the paper.

Competing interests E.W. is co-founder of Vescor Therapeutics. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0697-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0697-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to E.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mice. All animal care and treatments were carried out in compliance with Rutgers University Institutional Animal Care and Use Committee guidelines (IACUC). Mice for conditional whole-body deletion of *Atg7* (C57Bl/6J *Ubc-cre*^{ERT2/+}; *Atg7*^{lox/lox}) were engineered with floxed alleles of *Atg7* (*Atg7*^{lox/lox})² and a transgene expressing the TAM-regulated Cre recombinase fusion protein under the control of the ubiquitously expressed ubiquitin C promoter (*Ubc*)³¹, as previously described¹. Acute deletion of *Atg7* throughout the mouse is obtained after TAM injection¹. TAM (T5648, Sigma) was suspended at a concentration of 20 mg/ml, in a mixture of 98% sunflower seed oil and 2% ethanol and 250 µl per 25 g of body weight was injected intraperitoneally into 8–10-week-old male *Ubc-cre*^{ERT2/+}; *Atg7*^{+/+} or *Ubc-cre*^{ERT2/+}; *Atg7*^{lox/lox} mice once per day for four days to generate cohorts of *Atg7*-deleted (*Atg7*^{Δ/Δ}) and wild-type (*Atg7*^{+/+}) control host mice. To assess the consequence of acute deletion of *Atg7* on tumorigenesis of C57Bl/6J isogenic male tumour cells, one week after TAM treatment, YUMM 1.1 (1 × 10⁶ cells), 1.3 (2 × 10⁶ cells), 1.7 (0.1 × 10⁶ cells), 1.9 (1 × 10⁶ cells), 71.8 (1 × 10⁶ cells) or MB49 (0.25 × 10⁶ cells) cells were resuspended in 100 µl PBS and injected subcutaneously into the dorsal flanks of mice. Three weeks after cell injection, mice were killed and serum and tumours were collected. The maximal tumour volume (1,700 mm³) permitted by Rutgers University IACUC was never exceeded. For arginine supplementation, 1% arginine (A8094, Sigma) in drinking water was given to the mice a week before TAM and throughout the experiment.

Mice for conditional whole-body deletion of *Atg5* (C57Bl/6J *Ubc-cre*^{ERT2/+}; *Atg5*^{lox/lox}) were engineered with floxed alleles of *Atg5* (*Atg5*^{lox/lox})³² and a transgene expressing the TAM-regulated Cre recombinase fusion protein under the control of the *Ubc* promoter³¹. Acute deletion of *Atg5* throughout mice was obtained after TAM injection (250 µl of TAM per 25 g of body weight injected intraperitoneally into 8–10-week-old male *Ubc-cre*^{ERT2/+}; *Atg5*^{+/+} or *Ubc-cre*^{ERT2/+}; *Atg5*^{lox/lox} mice once a week for four weeks) to generate cohorts of *Atg5*-deleted (*Atg5*^{Δ/Δ}) and wild-type (*Atg5*^{+/+}) control host mice.

Liver-specific deletion of *Atg7* and *Atg5* was achieved by injecting an adeno-associated virus (AAV)–thyroxine binding globulin (TBG) promoter–Cre recombinase vector (AAV–TBG–iCre, Vector Biolabs) into *Atg7*^{lox/lox} and *Atg5*^{lox/lox} mice. An AAV–TBG promoter–GFP vector (AAV–TBG–GFP, Vector Biolabs) was injected into *Atg7*^{lox/lox} and *Atg5*^{lox/lox} mice as a control. 1.5 × 10¹¹ genome copies of either AAV–TBG–iCre or AAV–TBG–GFP vectors in 100 µl PBS were injected into the tail vein of 8–10-week-old male *Atg7*^{lox/lox} and *Atg5*^{lox/lox} mice to generate liver-specific *Atg7*^{Δ/Δ} or *Atg5*^{Δ/Δ} and *Atg7*^{+/+} or *Atg5*^{+/+} control mice, respectively. Three weeks post injection, YUMM 1.1 cells (1 × 10⁶ cells) were resuspended in 100 µl PBS and injected subcutaneously into the dorsal flanks of the liver-specific *Atg7*^{Δ/Δ} and *Atg7*^{+/+} control mice. Tumour growth was monitored daily. Tumour volume was calculated with the following formula: volume = π/6 × L × W × H. Three weeks after cell injection, mice were killed and liver, kidney, brain, serum and tumours were collected.

Cell lines. Cell culture. Cell lines were authenticated using whole-exome sequencing. YUMM 1.1, 1.3, 1.7 and 1.9 cells derived from *Braf*^{V600E/+} *Pten*^{−/−} *Cdkn2a*^{−/−} C57Bl/6J mouse melanomas were previously generated³³ and cultured in Dulbecco's minimum essential medium and Ham's F12 (DMEM-F12) (10-092-CV, Corning) supplemented with 10% fetal bovine serum (FBS) (F0926, Sigma) in a 5% CO₂ incubator at 37°C. The mouse lung cancer cell line 71.8 was previously derived from *p53*^{−/−} *Kras*^{G12D/+} mouse lung tumours¹¹ and the MB49 cell line³⁴ was provided by the Ratliff laboratory and cultured in Roswell Park Memorial Institute medium (RPMI) (11875-093, Gibco). Cells were tested negative for mycoplasma contamination.

Cell proliferation in arginine-deficient medium. YUMM 1.1, 1.3, 1.7, 1.9, 71.8 and MB49 cells were seeded at a density of 15,000 cells per well in 24-well plates. The following day, cells were washed with phosphate-buffered saline (PBS) (14190-144, Gibco) and cultured in arginine-free DMEM-F12 (DFL27, Caisson Labs) or arginine-free RPMI (R1780, Sigma) supplemented with 10% dialysed FBS (89986, Thermo Scientific) and an increasing percentage of arginine from 2.5 to 100%. Growth was assessed using an IncuCyte ZOOM with images of the proliferative cells recorded every 2 h for a total duration of 6 days.

Metabolite analysis by LC–MS. *Metabolite extraction for LC–MS.* Metabolites from 10-µl serum samples were first extracted with 40 µl of ice-cold methanol. The mixture was allowed to sit at −20°C for 20 min, and then centrifuged at 16,000g for 10 min at 4°C. Supernatants were transferred to clean tubes and pellets were extracted again with 200 µl 40:40:20 methanol:acetonitrile:H₂O. The mixture was allowed to sit on ice for 10 min, and then centrifuged at 16,000g for 10 min at 4°C. Supernatants were combined with the first extraction, resulting in roughly 240 µl of extract. Extracts were further processed with Phree Phospholipid Removal 1-ml Tube (Phenomenex), according to the manufacturer's instructions. The final extract was stored at −80°C until analysis by LC–MS. To extract metabolites from the tissues and tumours, samples (25 mg) were first pulverized using a Cryomill (Retsch) in liquid nitrogen at 25 Hz for 2 min. Extraction was performed by adding

−20°C 40:40:20 methanol:acetonitrile:water with 0.5% formic acid solution (500 µl) to the ground samples, followed by vortexing and centrifugation at 16,000g for 10 min at 4°C. The supernatants were transferred to clean tubes and the pellets were extracted again by repeating the previous step. The supernatant was then combined with the first extract. Then, 500 µl of extract was neutralized with 44 µl of 15% NH₄HCO₃ solution and centrifuged at 16,000g for 10 min at 4°C to remove protein precipitate. Then, 300 µl of supernatant was removed to clean tubes and stored at −80°C until analysis by LC–MS.

LC–MS analysis. LC–MS analysis of the extracted metabolites was performed on a Q Exactive PLUS hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific) coupled to hydrophilic interaction chromatography. The LC separation was performed on an UltiMate 3000 UHPLC system with an XBridge BEH Amide column (150 mm × 2.1 mm, 2.5 µm particle size, Waters) with the corresponding XP VanGuard Cartridge. The liquid chromatography used a gradient of solvent A (95%:5% H₂O:acetonitrile with 20 mM ammonium acetate, 20 mM ammonium hydroxide, pH 9.4), and solvent B (20%:80% H₂O:acetonitrile with 20 mM ammonium acetate, 20 mM ammonium hydroxide, pH 9.4). The gradient was 0 min, 100% B; 3 min, 100% B; 3.2 min, 90% B; 6.2 min, 90% B; 6.5 min, 80% B; 10.5 min, 80% B; 10.7 min, 70% B; 13.5 min, 70% B; 13.7 min, 45% B; 16 min, 45% B; 16.5 min, 100% B. The flow rate was 300 µl/min. Injection volume was 5 µl and column temperature 25°C. The mass spectrometry scans were in negative-ion mode with a resolution of 70,000 at *m/z* 200. The automatic gain control target was 3 × 10⁶ and the scan range was 75–1,000. To increase metabolome coverage, the samples were also analysed with a secondary LC–MS method, which involves two separate instrument platforms covering both positively charged and negatively charged metabolites. Negatively charged metabolites were analysed via reverse-phase ion-pairing chromatography coupled to an Exactive orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in negative ion mode with resolving power of 100,000 at *m/z* 200, scanning range being *m/z* 75–1,000. The liquid chromatography method has previously been described³⁵, using a Synergy Hydro-RP column (100 mm × 2 mm, 2.5 µm particle size, Phenomenex) with a flow rate of 200 µl/min. The liquid chromatography gradient was 0 min, 0% B; 2.5 min, 0% B; 5 min, 20% B; 7.5 min, 20% B; 13 min, 55% B; 15.5 min, 95% B; 18.5 min, 95% B; 19 min, 0% B; 25 min, 0% B. Solvent A is 97:3 water:methanol with 10 mM tributylamine and 15 mM acetic acid; solvent B is methanol. Positively charged metabolites were analysed on a Q Exactive Plus mass spectrometer coupled to Vanquish UHPLC system (Thermo Fisher Scientific). The mass spectrometer was operated in positive-ion mode with resolving power of 140,000 at *m/z* 200, scanning range being *m/z* 75–1,000. The liquid chromatography separation was achieved on an Agilent Poroshell 120 Bonus-RP column (150 × 2.1 mm, 2.7 µm particle size). The gradient was 0 min, 50 µl/min, 0.0% B; 6 min, 50 µl/min, 0% B; 12 min, 200 µl/min, 70% B; 14 min, 200 µl/min, 100% B; 18 min, 200 µl/min, 100% B; 19 min, 200 µl/min, 0% B; 24 min, 200 µl/min, 0% B; 25 min, 50 µl/min, 0% B. Solvent A is 10 mM ammonium acetate + 0.1% acetic acid in 98:2 water:acetonitrile and solvent B is acetonitrile³⁶. Metabolite features were extracted in MAVEN v.707³⁷ with the labelled isotope specified and a mass accuracy window of 5 p.p.m. For the ¹³C¹⁵N arginine infusions, the isotope natural abundance and impurity of labelled substrate was corrected using a matrix-based algorithm.

Labelled arginine infusion. For jugular vein catheterization, the procedure was modified from work previously described³⁸. In brief, *Atg7*^{Δ/Δ} and *Atg7*^{+/+} mice were anaesthetized using isoflurane carried by oxygen, followed by placement of a central venous catheter (polyurethane tubing, 1 F in OD) (SAI Infusion Technologies) into the right jugular vein. A minimal amount of blood was carefully withdrawn to verify the catheter patency. Afterwards, the saline solution in the catheter was replaced by heparin–glycerol catheter lock solution (SAI Infusion Technologies). The proximal end of the catheter was then tunnelled subcutaneously, exited between the shoulder blades and properly secured. A fully recovered surgical mouse was placed in a plastic harness (SAI Infusion Technologies), and the catheter was connected to an infusion pump (New Era Pump System) through a mouse tether and swivel system (Instech Laboratories). Arginine isotope tracer (¹³C₆¹⁵N₄, CNLM-539-H-PK, Cambridge Isotope Laboratories) was dissolved in sterile saline and infused at a rate of 3.5 nmol/g/min (0.1 µl/g/min) for 3 h. Infusion rate was determined using turnover flux calculations²¹. Mice were killed after infusion for serum, tumour, liver and kidney analysis by LC–MS. The isotope natural abundance and impurity of labelled substrate was corrected using a matrix-based algorithm. The construction of the purity matrix and C/N joint correction matrix is similar to AccuCor³⁹. For calculation of the circulating amino acid concentration, the average ion counts from the *Atg7*^{+/+} mice were normalized to the previously measured amino acid concentration⁴⁰. The amino acid concentrations in the *Atg7*^{Δ/Δ} mice were calculated proportionally.

Arginine activity assay. To follow conversion of arginine to ornithine, 15 µl serum samples were added to 5 µl of 9.7 mM MnCl₂, 5 µl of 360 mM pH 9.7 glycine and 5 µl of 300 µM ¹³C₆-arginine followed by incubation at 37°C for 0, 5, 20, 60 or 120 min. Then, 870 µl of 40:40:20 methanol:acetonitrile:water with 0.5% formic

acid solution were added to stop the reaction; the mixture was allowed to sit on ice for 10 min. The extract was neutralized with 40 μ l of 15% NH_4HCO_3 solution and centrifuged at 16,000g for 10 min at 4°C. Then, 500 μ l of supernatant was removed to clean tubes and stored at -80°C until analysis by LC–MS. The LC–MS analysis was performed on the Q Exactive PLUS mass spectrometer coupled to UltiMate 3000 UHPLC system with an XBridge BEH Amide column (150 mm \times 2.1 mm, 2.5 μ m particle size, Waters) with the corresponding XP VanGuard Cartridge. The liquid chromatography used a 6-min isocratic elution of 28% solvent A (95%:5% H_2O :acetonitrile with 20 mM ammonium acetate, 20 mM ammonium hydroxide, pH 9.4) and 72% solvent B (20%:80% H_2O :acetonitrile with 20 mM ammonium acetate, 20 mM ammonium hydroxide, pH 9.4). The flow rate was 300 μ l/min. Injection volume was 5 μ l and column temperature 25°C. The mass spectrometry scans were in negative-ion mode with a resolution of 70,000 at m/z 200. The automatic gain control target was 3×10^6 and the scan range was 75–1,000. Metabolite features were extracted in MAVEN v.707 with the labelled isotope specified and a mass accuracy window of 5 p.p.m.

Proteomic analysis by LC–MS. Technical duplicates of pooled serum samples from both $\text{Atg}^{7+/+}$ ($n = 5$) and $\text{Atg}^{7\Delta/\Delta}$ ($n = 5$) mice were processed in parallel. Two different methods were also used to reduce the amount of major serum proteins, to allow detection of rarer components: AlbuVoid (Biotech Support Group) was used to deplete albumin, and the Agilent multiple affinity removal spin cartridge mouse 3 system (Mars3) was used to remove albumin, IgG and transferrin following the manufacturer's protocol. Untreated or depleted sera were loaded onto NuPage 10% Bis-Tris Gel (Invitrogen), run a short distance into the gel, and proteins reduced, alkylated and digested with trypsin as described⁴¹. Digests were analysed by nano LC–MS/MS using a Dionex Ultimate 3000 RLC nano System interfaced with Q Exactive HF (ThermoFisher). Peptides were loaded onto a self-packed 100 μ m \times 2 cm trap (Magic C18AQ, 5 μ m 200 Å, Michrom Bio resources) and washed with buffer A (0.1% trifluoroacetic acid) for 5 min with a flow rate of 10 μ l/min. The trap was brought in-line with the analytical column (self-packed Magic C18AQ, 3 μ m 200 Å, 75 μ m \times 50 cm) and fractionated at 300 nl/min using a segmented linear gradient of 4–15% B in 30 min (A: 0.2% formic acid; B: 0.16% formic acid/80% acetonitrile), 15–25% B in 40 min, 25–50% in 44 min and 50–90% B in 11 min. Mass spectrometry data were acquired using a data-dependent acquisition procedure with each cycle consisting of a MS1 scan (resolution 120,000) followed MS/MS scans (HCD relative collision energy 27%, resolution 30,000) of the 20 most intense ions using a dynamic exclusion duration of 20 s. The raw data were converted into MASCOT generic format using Proteome Discover 2.1 (Thermo Fisher) and searched against the Ensemble mouse database and a database of common laboratory contaminants (<http://www.thegpm.org/crap/>) using a local implementation of the global proteome machine (GPM Fury)⁴². Peptide spectrum matches were assigned to genes using BioMart Ensembl tables. To estimate differential abundances of proteins, data from all LC–MS runs were combined (neat, AlbuVoid-depleted and Mars3-depleted for each of the four samples). For mouse proteins with 10 or more spectral counts, differential expression was estimated using the QLSpline option of the QuasiSeq package (<https://cran.r-project.org/web/packages/QuasiSeq/index.html>)⁴³. Data are presented as fold change (thresholded $\log_2(\text{Atg}^{7\Delta/\Delta}/\text{Atg}^{7+/+})$) with adjusted P values < 0.05 . P values were adjusted using Holm correction using the 'p.adjust' function in the base R package (<https://cran.r-project.org>). The raw mass spectrometry data have been deposited in the MassIVE repository, entry MSV000082879.

Enzymatic assays. Levels of nitric oxide in serum were determined with the nitric oxide assay kit (ab65328, Abcam).

Histology. Mouse tissues were fixed in 10% buffer formalin solution overnight and then transferred to 70% ethanol for paraffin-embedded sections. Tissue sections were deparaffinized, rehydrated and boiled for 45 min in 10 mM pH 6 citrate buffer. Slides were blocked in 10% goat serum for an hour and then incubated at 4°C overnight with primary antibody against Ki-67 (1:200, Ab15580, Abcam), active caspase-3 (1:300, 9661, Cell Signaling), CD3 (1:100, Ab16669, Abcam), CD4 (1:1,000, Ab183685, Abcam) and CD8 (1:100, 14-0808-82, Invitrogen). The following day, tissue sections were incubated with biotin-conjugated secondary antibody for 15 min (Vector Laboratories), 3% hydrogen peroxide for 5 min, horseradish peroxidase streptavidin for 15 min (SA-5704, Vector Laboratories) and developed by 3,3'-diaminobenzidine (Vector Laboratories) followed by haematoxylin staining (3536-16, Ricca). Sections were then dehydrated, mounted in Cytoseal 60 mounting medium (8310, Thermo Scientific) and analysed using Nikon Eclipse 80i microscope. For quantification of immunohistochemistry, at least 10 images containing a minimum of 100 cells were analysed at 60 \times magnification for each genotype.

Western blotting. Tissues and tumour samples were grounded in liquid nitrogen, lysed in Tris lysis buffer (50 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 0.1% NP40, 5 mM MgCl_2 , 10% glycerol), separated on 12.5% SDS–PAGE gel and then transferred on PVDF membrane (Millipore). Membranes were blocked with 5% non-fat milk for 1 h and probed overnight at 4°C with antibodies against ASS1 (1:1,000,

Ab170952, Abcam), ASL (1:500, sc-374353, Santa Cruz), OTC (1:500, sc-515791, Santa Cruz), ARG1 (1:500, sc-271430, Santa Cruz), ATG7 (1:2,000, A2856, Sigma), transferrin (1:1,000, sc-22597, Santa Cruz), ATG5 (1:1,500, Ab108327, Abcam) and β -actin (1:5,000, A1978, Sigma). Immunoreactive bands were detected using peroxidase-conjugated antibody (GE Healthcare) and enhanced chemiluminescence detection reagents (NEL105001EA, Perkin Elmer) and were analysed using the ChemiDoc XRS+ system (Biorad). Protein levels were quantified using the Image Laboratory v.6.0.1 software. Antibodies were validated with the use of positive and negative controls, following the manufacturer's protocol.

T cell depletion and flow cytometry. A week after TAM, and every 5 days, 200 μ g of CD4 (clone GK1.5; BE003-1, BioXCell) and CD8 (clone 2.43; BE0061, BioXCell) antibodies were injected intraperitoneally into $\text{Atg}^{7\Delta/\Delta}$ and $\text{Atg}^{7+/+}$ mice. Two days after the first antibody injection, YUMM 1.1 (1×10^6 cells) cells were resuspended in 100 ml PBS and injected subcutaneously into the dorsal flanks of the mice. Three weeks after cell injection, mice were killed and tumours and spleen were collected. Tumours were homogenized in PBS in a gentleMACS Octo Dissociator (Miltenyi Biotec), according to the manufacturer's protocol, and passed through a 70-mm cell restrainer. Spleens were ground with a rubber grinder through steel mesh, treated with ACK Lysis Buffer to remove erythrocytes and passed through a 70-mm cell restrainer. Nonspecific binding of antibodies to cell Fc receptors was blocked using 20 ml per 10^7 cells of FcR blocker (Miltenyi Biotec). Cell surface immunostaining was performed with the following antibodies (1:200): CD11c-PE-eFluor610 (clone N418, 61-0114-82), CD4-APC (clone GK1.5, 17-0041-82) CD3-AF700 (clone 17A2, 56-0032-82) and CD11b-APC-Cy7 (clone M1/70, A15390) (eBioscience); and CD45-FITC (clone 30-F11, 103107), MHC-II-BV605 (clone M5/114.15.2, 107639), Ly6G-BV650 (clone 1A8, 127641) and CD8-BV785 (clone 53.67, 100749) (BioLegend). Aqua Live/Dead (Invitrogen) was included to determine live cells. After staining of surface markers, cells were fixed and permeabilized using transcription factor staining kit and stained with FoxP3-eFluor450 (eBioscience). Cell staining data were acquired using a LSR-II flow cytometer (BD Biosciences, BD FACS Diva v2 software) and analysed with FlowJo v.10 software (Tree Star). Live lymphocytes were gated using forward scatter area (FSC-A) versus side scatter area (SSC-A), followed by FSC-A versus forward scatter height (FSC-H), SSC-A versus side scatter height (SSC-H) plots, forward scatter width (FSC-W) versus side scatter width (SSC-W), and Aqua Live/Dead. Populations were gated as follows: CD45 (percentage CD45⁺ of total live lymphocytes), CD3 (percentage CD3⁺ of CD11b⁺CD11c⁺CD45⁺), CD8 (percentage CD8⁺ of CD3), CD4 (percentage CD4⁺ of CD3), T_{reg} (percentage FoxP3⁺ of CD4), DC (percentage CD11c⁺ of MHC-II⁺CD45⁺) and MDSC (percentage Ly6G, CD11b⁺ of MHC-II⁺CD45⁺).

Antibodies for western blotting, flow cytometry and immunohistochemistry were validated with the use of positive and negative controls (gene knockouts and through the use of control tissues and cell lines), and following the manufacturer's protocol.

Statistics. All statistical analyses were performed with Graphpad Prism v.7 software using two-sided Student's t -tests, unless specified otherwise. The sample size was chosen in advance on the basis of common practice of the described experiment and is mentioned for each experiment. No statistical methods were used to pre-determine sample size. Each experiment was conducted with biological replicates and repeated multiple times. All attempts at replication were successful and no data were excluded. Mice were randomly allocated to experimental groups and the investigators were not blinded during the experiments and outcome assessment.

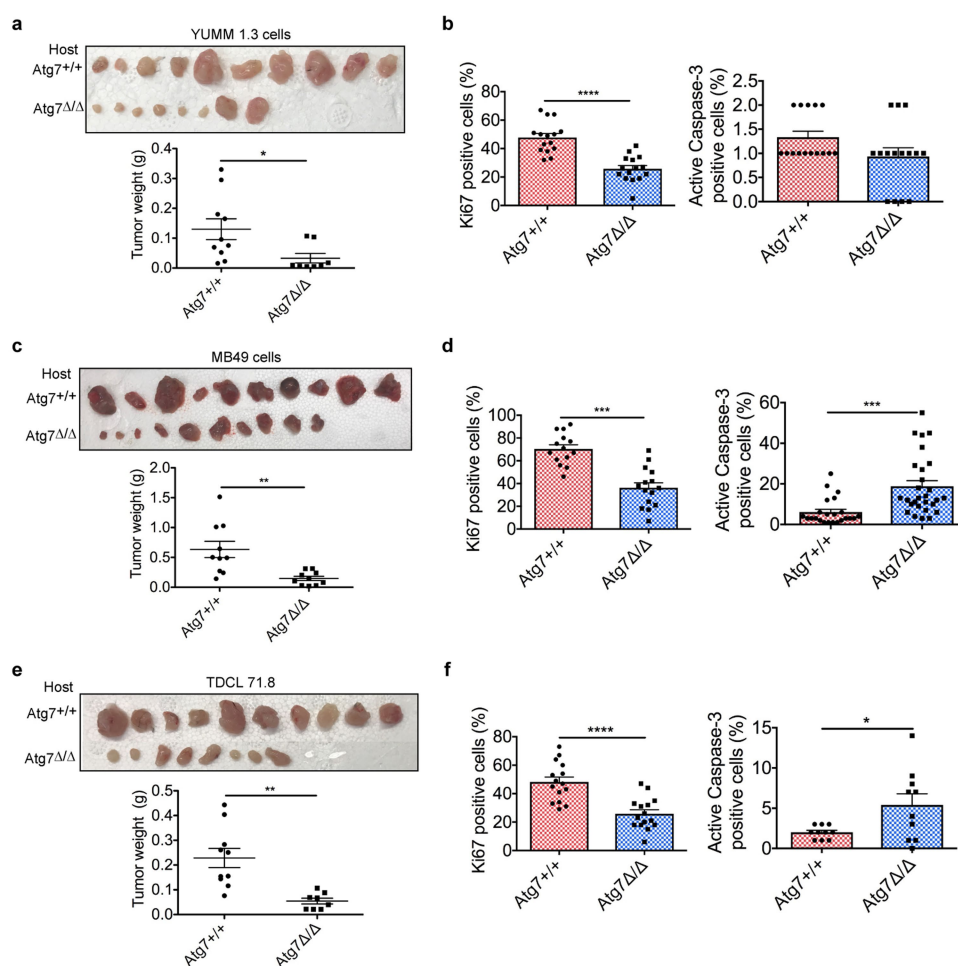
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data are available from the authors upon reasonable request. Source Data for Figs. 1d, 2b, 3d are provided with the paper. The raw mass spectrometry data have been deposited in the MassIVE repository, entry MSV000082879.

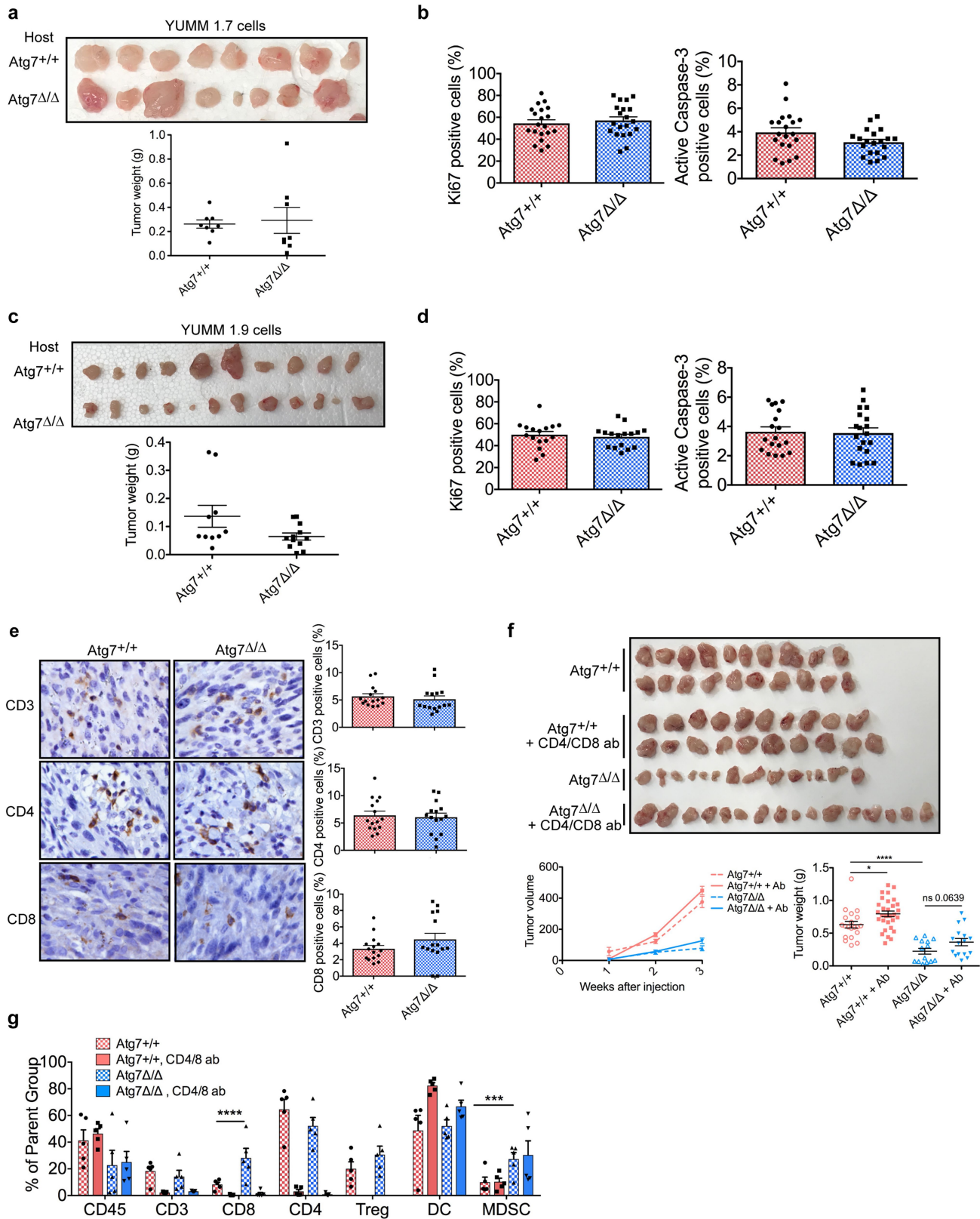
- Ruzankina, Y. et al. Deletion of the developmentally essential gene *ATR* in adult mice leads to age-related phenotypes and stem cell loss. *Cell Stem Cell* **1**, 113–126 (2007).
- Hara, T. et al. Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441**, 885–889 (2006).
- Meeth, K., Wang, J. X., Micevic, G., Damsky, W. & Bosenberg, M. W. The YUMM lines: a series of congenic mouse melanoma cell lines with defined genetic alterations. *Pigment Cell Melanoma Res.* **29**, 590–597 (2016).
- Summerhayes, I. C. & Franks, L. M. Effects of donor age on neoplastic transformation of adult mouse bladder epithelium in vitro. *J. Natl. Cancer Inst.* **62**, 1017–1023 (1979).
- Lu, W. et al. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal. Chem.* **82**, 3212–3221 (2010).
- Papazyan, R. et al. Physiological suppression of lipotoxic liver damage by complementary actions of HDAC3 and SCAP/SREBP. *Cell Metab.* **24**, 863–874 (2016).

37. Melamud, E., Vastag, L. & Rabinowitz, J. D. Metabolomic analysis and visualization engine for LC–MS data. *Anal. Chem.* **82**, 9818–9826 (2010).
38. Zhan, L. et al. Dysregulation of bile acid homeostasis in parenteral nutrition mouse model. *Am. J. Physiol. Gastrointest. Liver Physiol.* **310**, G93–G102 (2016).
39. Su, X., Lu, W. & Rabinowitz, J. D. Metabolite spectral accuracy on orbitraps. *Anal. Chem.* **89**, 5940–5948 (2017).
40. Sailer, M. et al. Increased plasma citrulline in mice marks diet-induced obesity and may predict the development of the metabolic syndrome. *PLoS ONE* **8**, e63950 (2013).
41. Sleat, D. E. et al. Mass spectrometry-based protein profiling to determine the cause of lysosomal storage diseases of unknown etiology. *Mol. Cell. Proteomics* **8**, 1708–1718 (2009).
42. Beavis, R. C. Using the global proteome machine for protein identification. *Methods Mol. Biol.* **328**, 217–228 (2006).
43. Lund, S. P., Nettleton, D., McCarthy, D. J. & Smyth, G. K. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat. Appl. Genet. Mol. Biol.* **11**, 1544–6115 (2012).



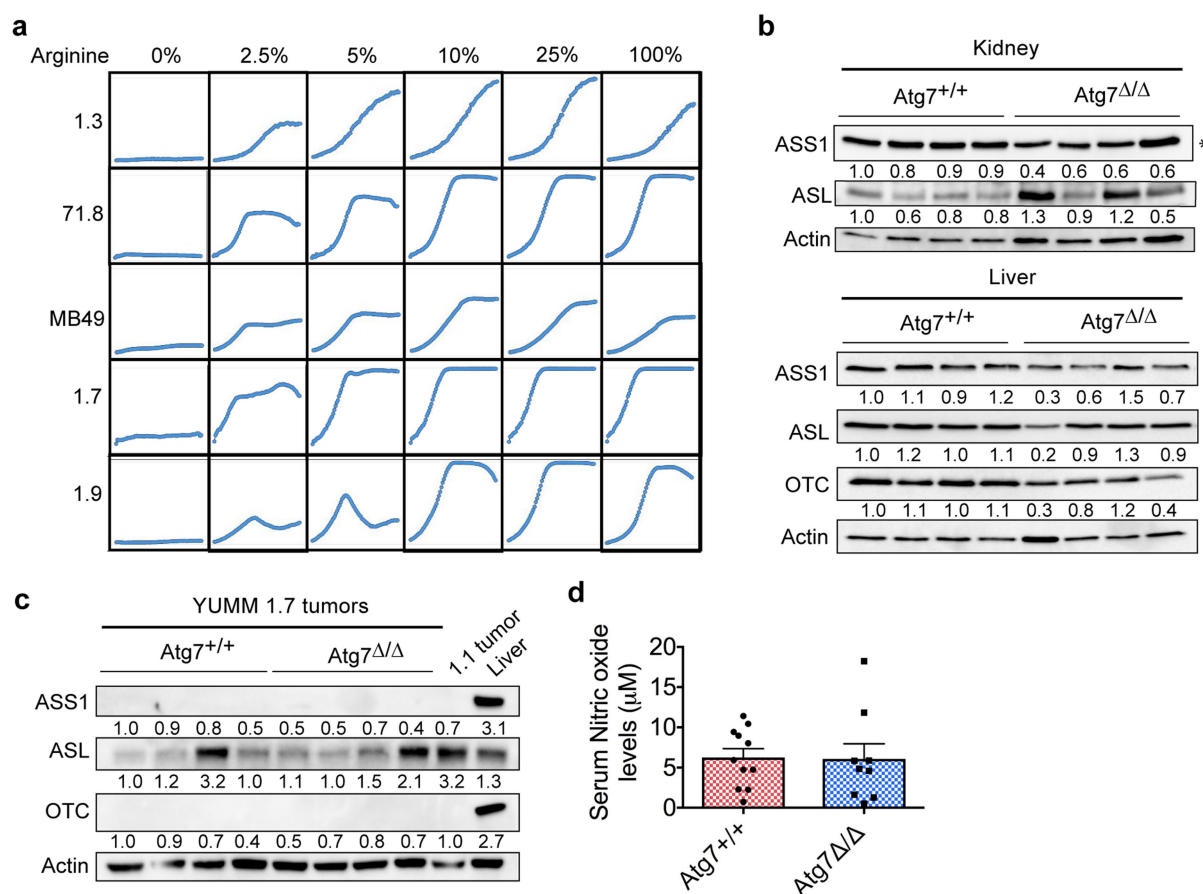
Extended Data Fig. 1 | Host autophagy promotes growth of different tumour cell types. a, c, e, Comparison of tumour weight between *Atg7*^{+/+} (*n* = 5) and *Atg7*^{Δ/Δ} (**a**, *n* = 4; **c**, *n* = 5; **e**, *n* = 4) hosts after injection of 1.3 (**a**), MB49 (**c**) or 71.8 (**e**) cells. Data are mean ± s.e.m. **P* < 0.05,

P* < 0.01. **b, d, f,** Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ cells in tumours from *Atg7*^{+/+} and *Atg7*^{Δ/Δ} hosts. Data are mean ± s.e.m. **P* < 0.05, *P* < 0.001, *****P* < 0.0001.



Extended Data Fig. 2 | Immune response is not involved in decreased tumour growth observed in *Atg7*^{Δ/Δ} hosts. **a, c,** Comparison of tumour weight between *Atg7*^{+/+} ($n = 5$) and *Atg7*^{Δ/Δ} ($n = 5$; $n = 6$) hosts after injection of 1.7 (**a**) or 1.9 (**c**) cells. Data are mean \pm s.e.m. **b, d,** Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ in 1.7 (**b**) and 1.9 (**d**) tumours from *Atg7*^{+/+} and *Atg7*^{Δ/Δ} hosts. Data are mean \pm s.e.m. **e,** Representative immunohistochemistry images and quantification of CD3⁺, CD4⁺ and CD8⁺ cells in tumours from *Atg7*^{+/+} and *Atg7*^{Δ/Δ} hosts. Data are mean \pm s.e.m. **f,** Comparison of

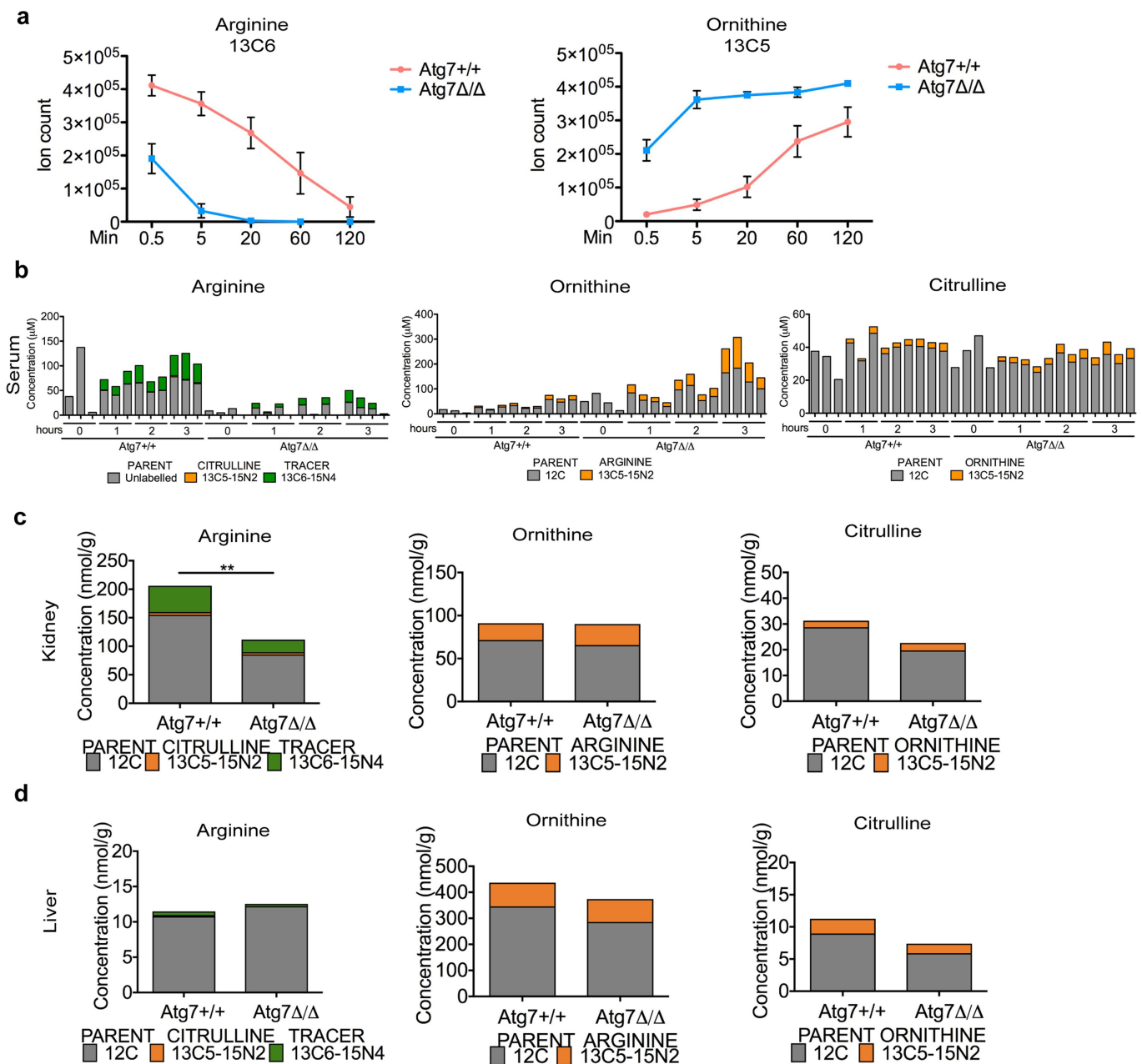
tumour volume and weight between *Atg7*^{+/+} ($n = 10$), *Atg7*^{+/+} + CD4 and CD8 antibody depletion ($n = 15$), *Atg7*^{Δ/Δ} ($n = 7$) and *Atg7*^{Δ/Δ} + CD4 and CD8 antibody depletion ($n = 8$) hosts. Data are mean \pm s.e.m. * $P < 0.05$, **** $P < 0.0001$. **g,** Fold change in immune components between *Atg7*^{+/+} and *Atg7*^{Δ/Δ}, with or without antibody depletion ($n = 5$ each). T_{reg}, T regulatory cells; DC, dendritic cells; MDSC, myeloid-derived suppressor cells. Data are mean \pm s.e.m. *** $P < 0.001$, **** $P < 0.0001$, by two-way ANOVA test.



Extended Data Fig. 3 | Tumour cells are arginine auxotrophs.

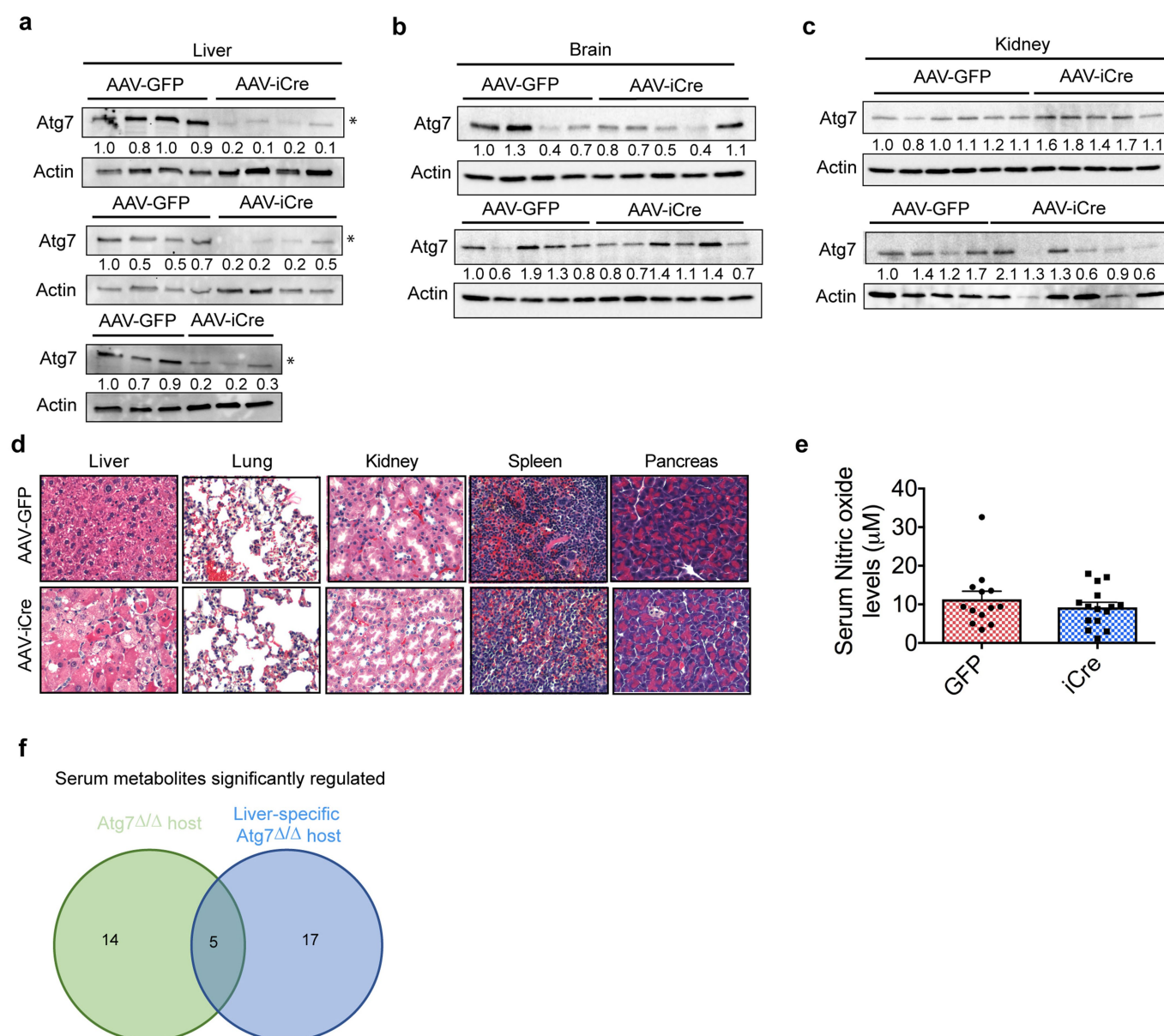
a, YUMM 1.3, 71.8, MB49 and YUMM 1.7, 1.9 proliferation in vitro, in medium containing different percentages of arginine. Cell density was measured every 2 h using the IncuCyte. Data are representative of three independent experiments performed in duplicate. **b**, Western blotting showing expression of ASS1, ASL and OTC in kidneys and livers from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts. **p* < 0.05 compared to *Atg7^{+/+}* hosts. Data are representative of three independent experiments. Actin was used as a

loading (kidney ASL and liver OTC) and processing (kidney ASS1, liver ASS1 and ASL) control. **c**, Western blotting showing expression of ASS1, ASL and OTC in YUMM 1.7 tumours from *Atg7^{+/+}* and *Atg7^{Δ/Δ}* hosts. Data are representative of two independent experiments. Actin was used as a loading (OTC) and processing (ASS1 and ASL) control. **d**, Analysis of levels of nitric oxide in serum in *Atg7^{+/+}* (*n* = 11) and *Atg7^{Δ/Δ}* (*n* = 9) hosts. Data are mean ± s.e.m.



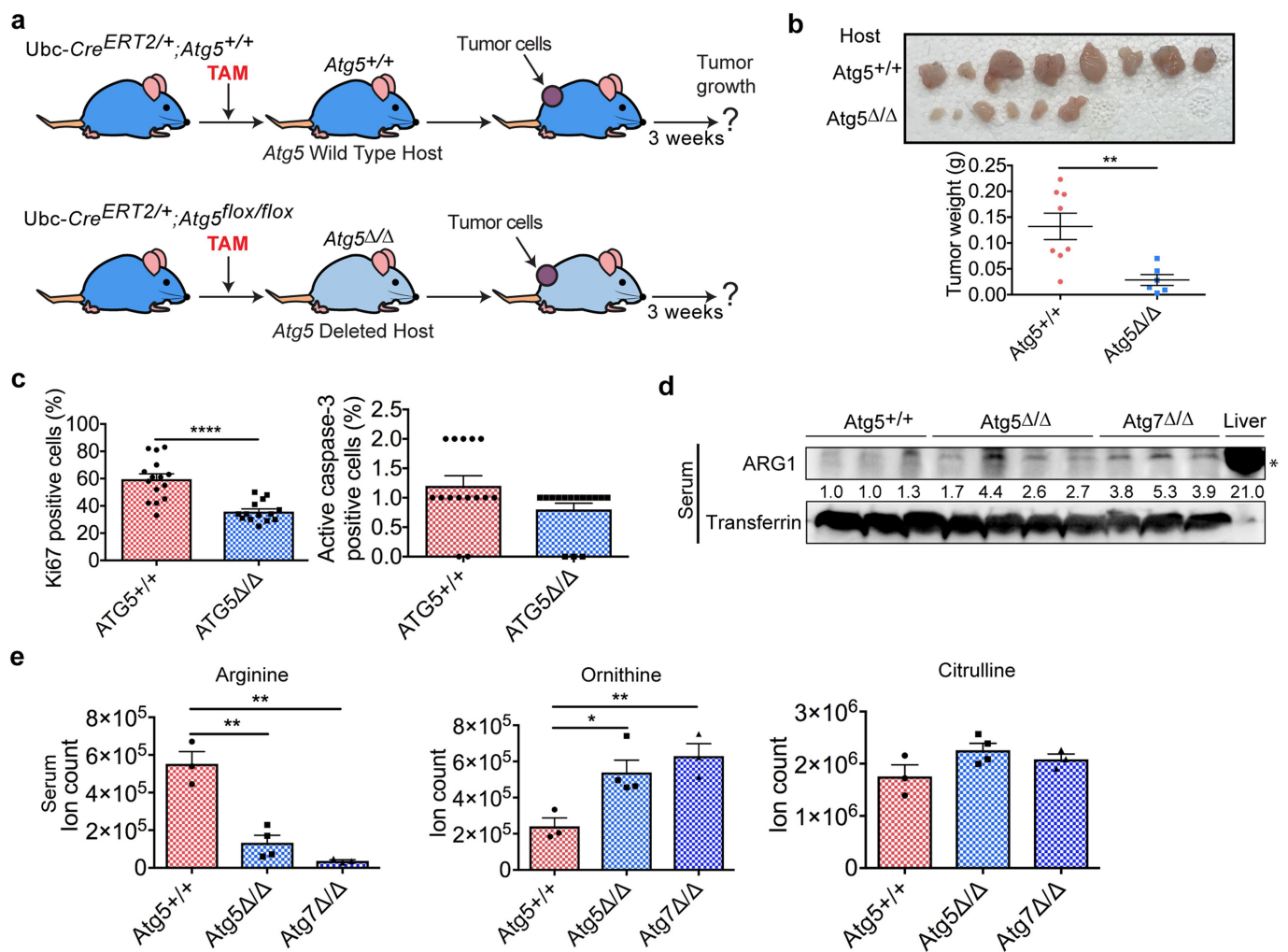
Extended Data Fig. 4 | Atg7 deletion increases serum arginine degradation but does not modify arginine metabolism in kidney and liver. **a**, Serum $^{13}\text{C}_6$ -arginine and $^{13}\text{C}_5$ -ornithine in Atg7^{+/+} and Atg7 Δ/Δ hosts ($n = 3$ each) over time. Data are mean \pm s.e.m. **b**, Concentration (in μM) of arginine, citrulline and ornithine in serum from Atg7^{+/+}

($n = 3$) and Atg7 Δ/Δ hosts ($n = 4$), after infusion with $^{13}\text{C}_6^{15}\text{N}_4$ -arginine. **c**, **d**, Concentration (in nmol g^{-1}) of arginine, citrulline and ornithine in kidneys (**c**) and livers (**d**) from Atg7^{+/+} and Atg7 Δ/Δ hosts ($n = 2$ each) after infusion with $^{13}\text{C}_6^{15}\text{N}_4$ -arginine. Data are mean. $**P < 0.01$ by two-way ANOVA test.



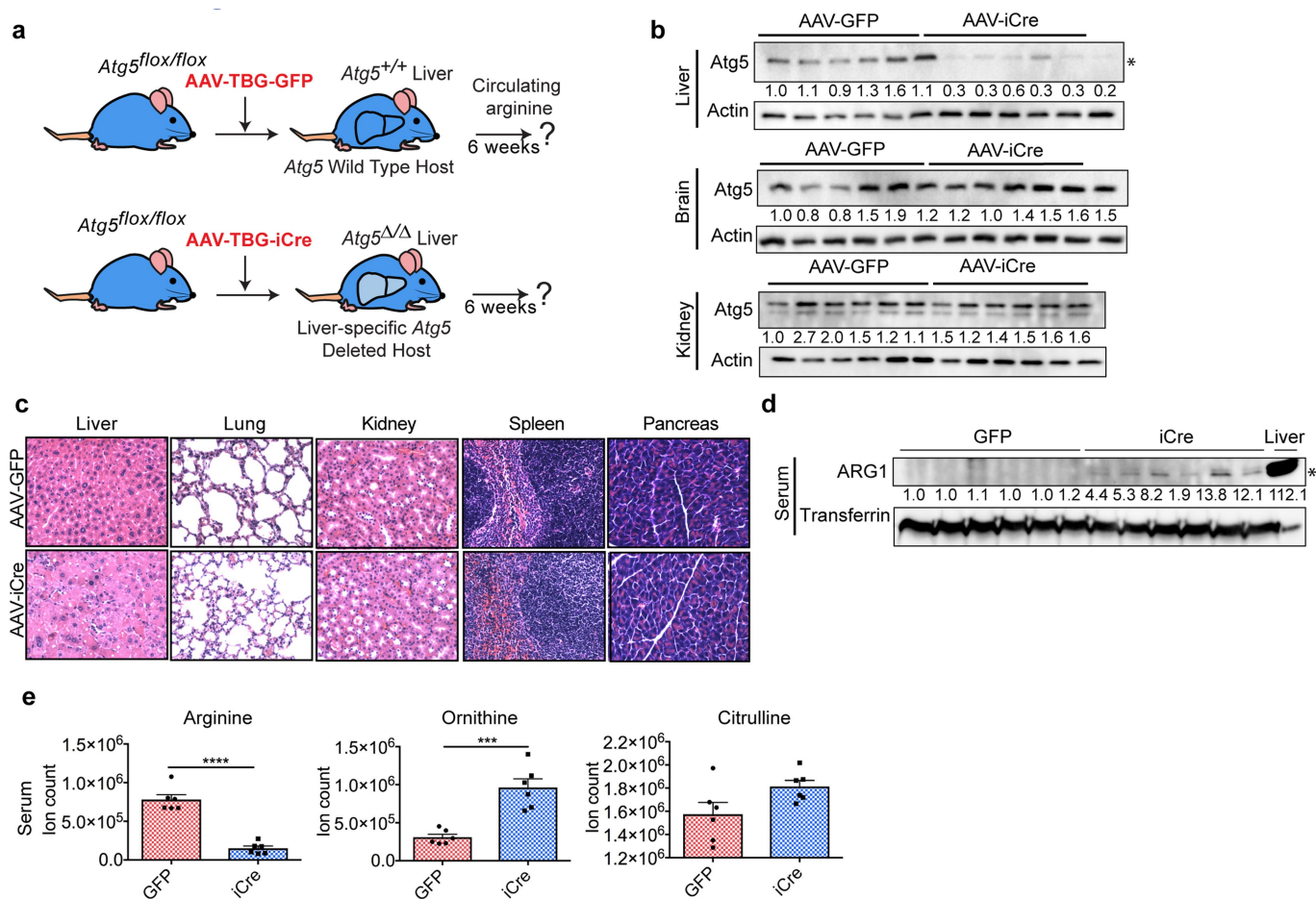
Extended Data Fig. 5 | Liver-specific deletion of *Atg7* leads to liver-cell enlargement without affecting other tissues. **a–c**, Western blotting showing expression of *Atg7* in livers ($n = 11$ each) (**a**), brains ($n = 9$ and 11 , respectively) (**b**) and kidneys ($n = 10$ each) (**c**) from *Atg7* $^{+/+}$ hosts and hosts with liver-specific deletion of *Atg7*. * $P < 0.05$ compared to *Atg7* $^{+/+}$ hosts. Data are representative of two independent experiments. Actin was used as a loading control. **d**, Representative haematoxylin and eosin

tissue staining from *Atg7* $^{+/+}$ hosts and hosts with liver-specific deletion of *Atg7*. Images are representative of two independent experiments. **e**, Analysis of levels of nitric oxide in serum, in *Atg7* $^{+/+}$ hosts ($n = 13$) and hosts with liver-specific deletion of *Atg7* ($n = 15$). Data are mean \pm s.e.m. **f**, Comparison of serum metabolites that are significantly regulated in *Atg7* Δ/Δ hosts and hosts with liver-specific deletion of *Atg7* ($n = 17$ each, $P < 0.05$).



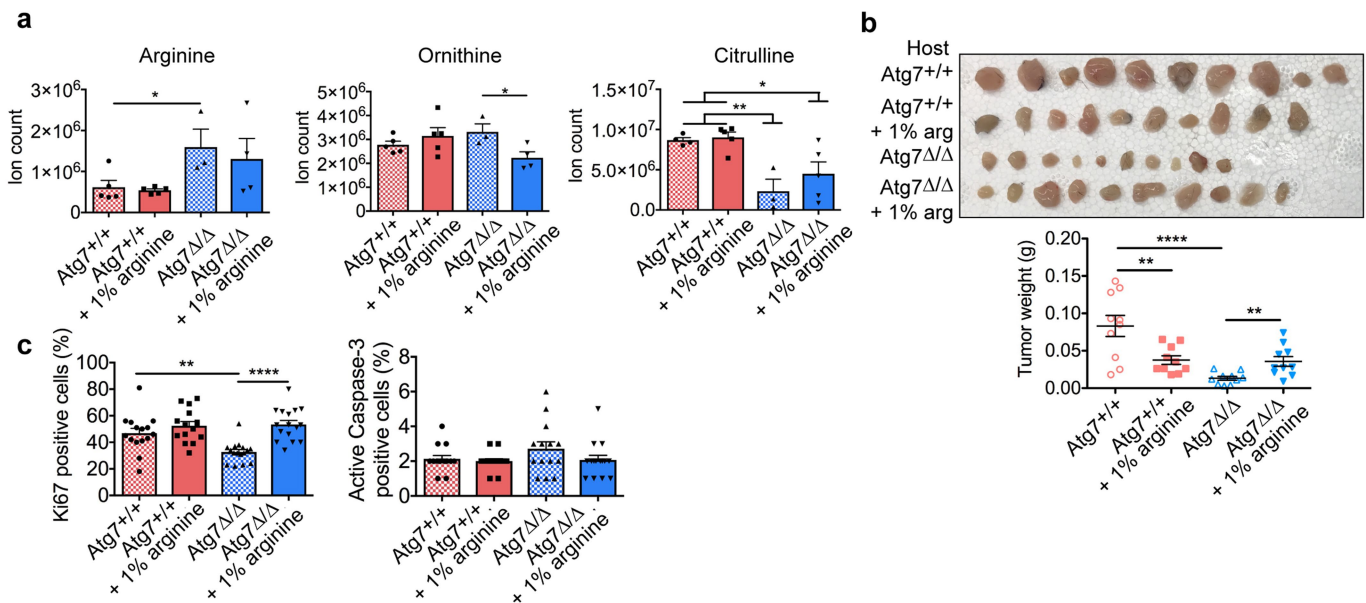
Extended Data Fig. 6 | Atg5 deletion increased serum ARG1, decreased serum arginine and tumour growth. **a**, Experimental design to induce host mice with conditional whole-body deletion of Atg5 (Atg5^{Δ/Δ}) and wild-type controls (Atg5^{+/+}) with which to assess tumour growth. Ubc-cre^{ERT2/+};Atg5^{+/+} and Ubc-cre^{ERT2/+};Atg5^{flox/flox} mice were injected with TAM at 8 to 10 weeks of age to delete Atg5 and create Atg5^{+/+} and Atg5^{Δ/Δ} hosts. Mice were then injected subcutaneously with tumour cells and tumour growth was monitored over three weeks. **b**, Comparison of tumour weight between Atg5^{+/+} ($n = 4$) and Atg5^{Δ/Δ} ($n = 3$) hosts. Data

are mean \pm s.e.m. $**P < 0.01$. **c**, Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ cells in tumours from Atg5^{+/+} and Atg5^{Δ/Δ} hosts. Data are mean \pm s.e.m. $****P < 0.0001$. **d**, Western blotting showing expression of ARG1 in serum from Atg5^{+/+} ($n = 3$), Atg5^{Δ/Δ} ($n = 4$) and Atg7^{Δ/Δ} ($n = 3$) hosts. $*P < 0.05$ compared to Atg5^{+/+} hosts. Transferrin was used as a loading control. **e**, Levels of arginine, ornithine and citrulline in serum in Atg5^{+/+} ($n = 4$) and Atg5^{Δ/Δ} ($n = 3$) hosts, obtained by LC-MS. Data are mean \pm s.e.m. $*P < 0.05$, $**P < 0.01$.



Extended Data Fig. 7 | Liver-specific *Atg5*-deleted hosts present liver-cell enlargement, increased serum ARG1 and decreased serum arginine. **a**, Experimental design to induce liver-specific deletion of *Atg5*. *Atg5^{flox/flox}* mice were injected in the tail vein with AAV-TBG-GFP or AAV-TBG-iCre at 8 to 10 weeks of age to delete *Atg5* in the liver and create *Atg5^{+/+}* hosts and hosts with liver-specific deletion of *Atg5*, respectively. **b**, Western blotting showing expression of *Atg5* in the livers, brains and kidneys of *Atg5^{+/+}* hosts and hosts with liver-specific deletion of *Atg5* ($n = 6$ each). * $P < 0.05$ compared to *Atg5^{+/+}* hosts. Actin was

used as a loading control **c**, Haematoxylin and eosin tissue staining from *Atg5^{+/+}* hosts and hosts with liver-specific deletion of *Atg5* ($n = 6$ each). **d**, Western blotting showing expression of ARG1 in serum from *Atg5^{+/+}* hosts and hosts with liver-specific deletion of *Atg5* ($n = 6$ each). * $P < 0.05$ compared to *Atg5^{+/+}* hosts. Transferrin was used as a loading control. **e**, Levels of arginine, ornithine and citrulline in serum in *Atg5^{+/+}* hosts and hosts with liver-specific deletion of *Atg5* ($n = 6$ each), obtained by LC-MS. Data are mean \pm s.e.m. *** $P < 0.001$, **** $P < 0.0001$.



Extended Data Fig. 8 | Dietary arginine supplementation rescues YUMM 1.3 tumour growth in *Atg7*^{Δ/Δ} hosts. **a**, Serum arginine, ornithine and citrulline in *Atg7*^{+/+} ($n = 5$), *Atg7*^{+/+} + 1% arginine ($n = 5$), *Atg7*^{Δ/Δ} ($n = 6$) and *Atg7*^{Δ/Δ} + 1% arginine ($n = 6$) hosts, obtained by LC-MS. Data are mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$. **b**, Comparison of YUMM 1.3 tumour weight between *Atg7*^{+/+} and *Atg7*^{Δ/Δ} ($n = 5$ each)

hosts, with or without arginine supplementation. Data are mean \pm s.e.m. ** $P < 0.01$, **** $P < 0.0001$. **c**, Immunohistochemistry quantification of Ki-67⁺ and active caspase-3⁺ cells in tumours from *Atg7*^{+/+} and *Atg7*^{Δ/Δ} hosts, with or without arginine supplementation. Data are mean \pm s.e.m. ** $P < 0.01$, **** $P < 0.0001$.

POLAR-guided signalling complex assembly and localization drive asymmetric cell division

Anaxi Houbaert^{1,2}, Cheng Zhang^{1,2}, Manish Tiwari^{1,2}, Kun Wang^{1,2,3}, Alberto de Marcos Serrano⁴, Daniel V. Savatin^{1,2}, Mounashree J. Urs^{1,2}, Miroslava K. Zhiponova^{1,2,6}, Gustavo E. Gudesblat^{1,2,7}, Isabelle Vanhoutte^{1,2}, Dominique Eeckhout^{1,2}, Sjeef Boeren⁵, Mansour Karimi^{1,2}, Camilla Betti^{1,2,8}, Thomas Jacobs^{1,2}, Carmen Fenoll⁴, Montaña Mena⁴, Sacco de Vries⁵, Geert De Jaeger^{1,2} & Eugenia Russinova^{1,2*}

Stomatal cell lineage is an archetypal example of asymmetric cell division (ACD), which is necessary for plant survival^{1–4}. In *Arabidopsis thaliana*, the GLYCOGEN SYNTHASE KINASE3 (GSK3)/SHAGGY-like kinase BRASSINOSTEROID INSENSITIVE 2 (BIN2) phosphorylates both the mitogen-activated protein kinase (MAPK) signalling module^{5,6} and its downstream target, the transcription factor SPEECHLESS (SPCH)⁷, to promote and restrict ACDs, respectively, in the same stomatal lineage cell. However, the mechanisms that balance these mutually exclusive activities remain unclear. Here we identify the plant-specific protein POLAR as a stomatal lineage scaffold for a subset of GSK3-like kinases that confines them to the cytosol and subsequently transiently polarizes them within the cell, together with BREAKING OF ASYMMETRY IN THE STOMATAL LINEAGE (BASL), before ACD. As a result, MAPK signalling is attenuated, enabling SPCH to drive ACD in the nucleus. Moreover, POLAR turnover requires phosphorylation on specific residues, mediated by GSK3. Our study reveals a mechanism by which the scaffolding protein POLAR ensures GSK3 substrate specificity, and could serve as a paradigm for understanding regulation of GSK3 in plants.

In *A. thaliana*, the stomatal lineage is initiated by an ACD of a committed meristemoid mother cell (MMC) that generates the meristemoid, and a larger stomatal lineage ground cell (SLGC). Meristemoids either differentiate into guard mother cells (GMCs), which can divide symmetrically to form a stoma, or undergo several rounds of amplifying ACDs to generate SLGCs. SLGCs can give rise to pavement cells or new satellite meristemoids through spacing ACDs^{1–4}. The transcription factor SPCH is required for these divisions, and its activity is negatively regulated through phosphorylation by the MAPK module^{8–10}. Prior to each ACD, the plant-specific protein BASL first localizes in the nucleus and, subsequently, polarizes to the distal site of the newly formed division plane^{11,12}. BASL interacts with the MAPK module and its polarization leads to elevated nuclear MAPK signalling and a reduced SPCH abundance in one of the two daughter cells¹³. After division, the BASL polarity is inherited by the SLGC only, enabling it to exit from the stomatal lineage¹¹.

BIN2 activity is regulated by brassinosteroid signalling, and BIN2 regulates stomatal development through inhibition of the MAPK signalling module and of SPCH^{5–7}. The *Arabidopsis* genome encodes ten SHAGGY-like kinases¹⁴ (ATSKs), of which at least seven function redundantly as negative regulators of brassinosteroid signalling^{15–19}. To discover novel proteins that regulate the substrate specificity of BIN2 during stomatal development, we identified BIN2-interacting proteins in the stomatal lineage by expressing GSgreen-tagged²⁰ BIN2 driven by the *TOO MANY MOUTHS* (*TMM*) promoter²¹ in *bin2-3 Arabidopsis* plants (*TMM:BIN2-GSgreen;bin2-3*) followed by immunoprecipitation

coupled to mass spectrometry (IP–MS). Enrichment of POLAR peptides (encoded by *At4g31805*) was detected (Supplementary Table 1). The reverse IP–MS analysis, using *POLAR:POLAR-GFP;Col-0*⁴ seedlings, identified BIN2 and seven other SHAGGY-like kinases as putative POLAR-interacting proteins (Supplementary Table 1), from which BIN2 and ATSK12 were selected for further studies. Co-immunoprecipitation experiments in *Arabidopsis* and tobacco (*Nicotiana benthamiana*) confirmed that POLAR co-purifies with BIN2 and ATSK12 (Fig. 1a, b). Because POLAR is likely to co-localize with BASL during ACD⁴, we investigated whether BIN2, ATSK12 and POLAR form a complex with BASL. As expected, POLAR co-purified with BIN2, ATSK12 and BASL, and both BIN2 and ATSK12 co-purified with BASL (Fig. 1b). The probable direct interactions between BIN2, ATSK12, POLAR and BASL were confirmed by ratiometric bimolecular fluorescence complementation (rBiFC) (Fig. 1c, d).

Next, we examined whether transient co-expression of BASL with either BIN2, ATSK12 or POLAR in tobacco leaf epidermis would trigger spontaneous polarization of these proteins, as shown for the MAPK kinase YODA (YDA)¹². Whereas overexpression of BASL alone resulted in its expression-level-dependent polarization, neither BIN2, ATSK12 nor POLAR was polarized when overexpressed on its own. BASL overexpression strongly polarized POLAR, but only weakly polarized BIN2 and ATSK12. However, BASL overexpression noticeably increased the polarization of BIN2 and POLAR when the three proteins were co-expressed (Extended Data Fig. 1). By contrast, co-expression of POLAR with BIN2 or ATSK12 did not trigger polarization of either protein. As reported for YDA¹², BASL was sufficient to polarize POLAR, whereas POLAR was required for the strong polarization of BIN2 and ATSK12 in the presence of BASL. In addition to changes in polarization, co-expression of POLAR with either BIN2 or its homologues resulted in nuclear exclusion of all nucleus-localized ATSKs, except ATSK32 (also known as ATSK0), whereas co-expression of BASL and BIN2 led to nuclear exclusion of BASL (Extended Data Figs. 2, 3). Moreover, the kinase activity of BIN2 was required for nuclear exclusion of BASL, but not for that of BIN2 in the presence of POLAR.

To study the function of BIN2, ATSK12, POLAR and BASL as a complex, we examined their expression pattern and subcellular localization in the abaxial epidermis of cotyledons of *Arabidopsis* plants expressing BIN2 (*BIN2:BIN2-GFP;Col-0*), ATSK12 (*ATSK12:ATSK12-GFP;atsk12*) (Extended Data Fig. 4a, b), POLAR (*POLAR:POLAR-GFP;polar-2*) and BASL (*BASL:GFP-BASL;basl-2*) at endogenous levels. BASL, POLAR and ATSK12 were expressed in protodermal cells in seedlings 1 day post-germination (d.p.g.), whereas BIN2 was undetectable (Extended Data Fig. 5a–d). BIN2 was detected at 2 d.p.g. throughout the epidermis and its expression increased with epidermal

¹Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. ²Center for Plant Systems Biology, VIB, Ghent, Belgium. ³College of Life Sciences, Wuhan University, Wuhan, China. ⁴Facultad de Ciencias Ambientales y Bioquímica, Universidad de Castilla-La Mancha, Toledo, Spain. ⁵Laboratory of Biochemistry, Wageningen University, Wageningen, The Netherlands. ⁶Present address: Department of Plant Physiology, Biological Faculty, University of Sofia, Sofia, Bulgaria. ⁷Present address: Instituto de Biodiversidad y Biología Experimental y Aplicada, Departamento de Fisiología, Biología Molecular y Celular, Universidad de Buenos Aires, Buenos Aires, Argentina. ⁸Present address: Department of Medicine, University of Perugia, Perugia, Italy. *e-mail: eurus@psb.vib-ugent.be

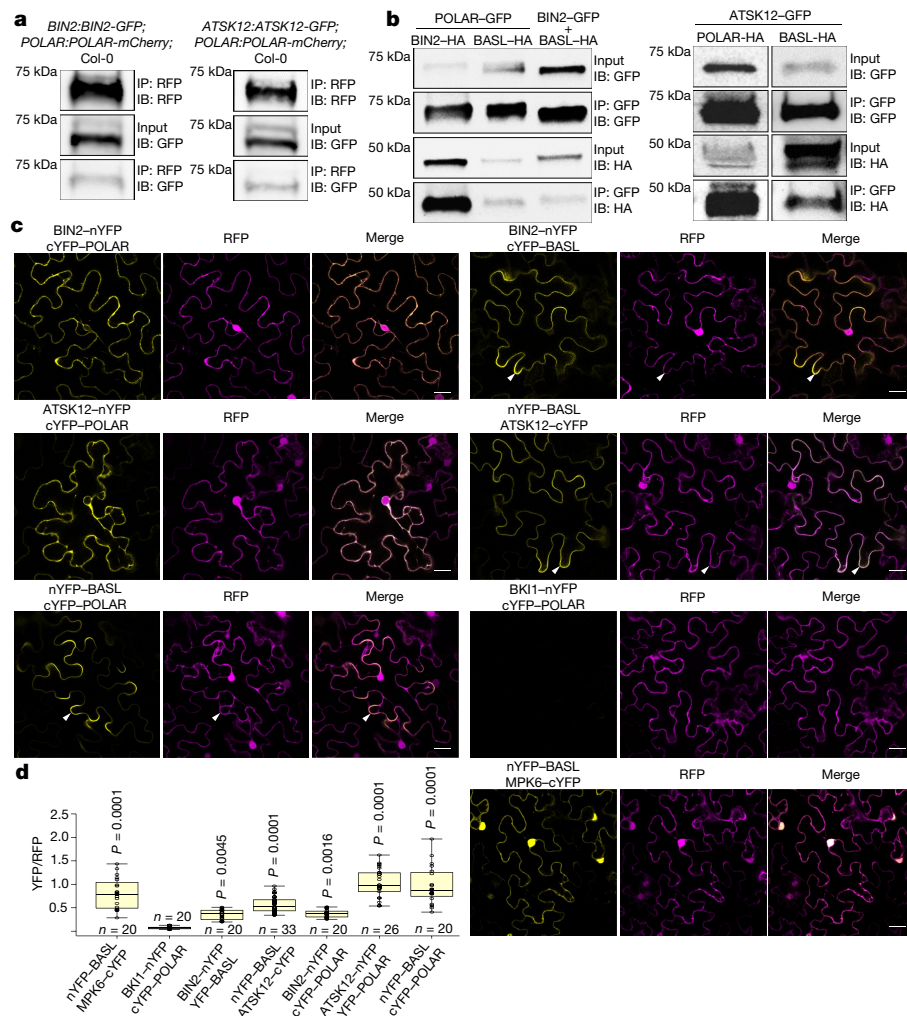


Fig. 1 | BIN2, ATSK12, BASL and POLAR form a complex. **a**, BIN2 and ATSK12 co-immunoprecipitated with POLAR in *Arabidopsis* seedlings at 3 d.p.g. POLAR-mCherry was not detected in the input. IP, immunoprecipitation; IB, immunoblot. **b**, BIN2, ATSK12 and POLAR formed complexes with BASL when co-expressed in tobacco. For blot source data for **a** and **b**, see Supplementary Fig. 1. **c**, Interaction between BIN2 or ATSK12 with either POLAR or BASL and between POLAR and BASL in tobacco leaf epidermis detected with rBiFC. The interactions between POLAR and BRI1 KINASE INHIBITOR1 (BK11) and between

BASL and MPK6 were used as negative and positive controls, respectively. White arrowheads indicate polarity. **d**, Quantification of the rBiFC in **c**. n , number of cells from three biologically independent leaves. Box plots throughout show the first and third quartiles, split by the median (line) and mean (cross), with whiskers extending $1.5 \times$ interquartile range beyond the box. One-way ANOVA with Tukey's post hoc test compared to the BK11-POLAR control. All experiments were repeated independently three times. Scale bars, 20 μ m.

cell maturation (Extended Data Fig. 5a). BIN2 appeared more abundant and more polarized at the cell cortex of stomatal lineage cells with ACD potential, and also displayed a nuclear localization. ATSK12 had a localization pattern similar to that of BIN2, except that it did not exhibit any nuclear signal in the epidermis when expressed at physiological levels (Extended Data Figs. 5b, 6a). Increased protein abundance and polarization in ACD precursors were observed for all examined BIN2 homologues (Extended Data Fig. 6a). We further performed time-lapse imaging and subdivided the ACD precursors (meristemoids and SLGCs) into three categories on the basis of BASL localization and cell morphology (Fig. 2a): ACD precursors with non-polarized BASL, ACD precursors with polarized BASL, and post-ACD cells corresponding to the newly formed meristemoid and SLGC (Fig. 2b). As previously reported¹¹, in ACD precursors, BASL first localized to the nucleus, and subsequently to the cell cortex immediately before ACD. Following ACD, the polarity was only inherited in SLGCs, and was enhanced from 30 to 60 min after ACD (Fig. 2b, c, Supplementary Video 1). In the same ACD precursors, POLAR, BIN2 and ATSK12 localized evenly throughout the plasma membrane and their association with the plasma membrane and polarization increased immediately before ACD. However, in contrast to BASL, following ACD, POLAR, BIN2

and ATSK12 exhibited decreased plasma membrane association, and polarity of POLAR remained in the SLGC, whereas polarity of BIN2 and ATSK12 was less evident (Fig. 2b, c, Supplementary Videos 2–4). POLAR displayed similar behaviour in asymmetrically dividing meristemoids and SLGCs (Fig. 2d). Notably, during ACD, both BIN2 and ATSK12 relocated to well-defined foci, resembling the spindle localization of mammalian GSK3s²² (Extended Data Fig. 5a, b). Furthermore, increasing the expression of POLAR, when co-expressed with either BIN2, ATSK12 or their homologues, led to an enhanced plasma membrane association and polarization of these proteins in the ACD precursors (Fig. 2b, c, Extended Data Figs. 5e, f, 6b, Supplementary Video 5). By contrast, loss-of-function of both POLAR and its close homologue POLAR-like1 (PL1)^{4,23} (encoded by *At5g10890*) (Extended Data Fig. 7) decreased the plasma membrane abundance and polarization of BIN2 in ACD precursors and led to a more pronounced nuclear BIN2 signal in the meristemoids (Fig. 2b, c, Extended Data Fig. 5e–h, Supplementary Video 6). Of note, increasing the expression of BASL affected neither BIN2 abundance at the plasma membrane nor its polarization (Extended Data Fig. 6c), further supporting a scaffolding function for POLAR. In summary, the transient polarization of BIN2 and ATSK12 during ACD is mediated by POLAR.

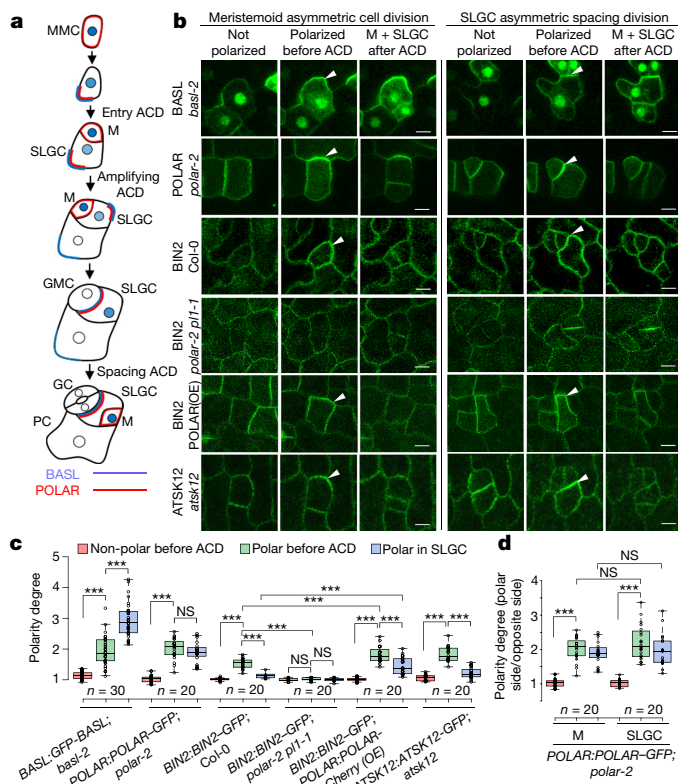


Fig. 2 | BIN2 and ATSK12 are polarized during stomatal ACD in *Arabidopsis*. **a**, Schematic representation of the localization of the BASL and POLAR proteins during ACD in the stomatal lineage. Blue, BASL; red, POLAR; M, meristemoid; PC, pavement cell; GC, guard cell. **b**, Localization of BASL, POLAR, BIN2 and ATSK12 during ACD in meristemoids or SLGCs at 2 d.p.g. **c**, Polarization of the genotypes in **b**. **d**, Polarization of POLAR in asymmetrically dividing meristemoids compared to SLGCs. *n*, number of cells from three biologically independent cotyledons (**c**, **d**). One-way ANOVA with Tukey's post hoc test. ****P* < 0.0005; NS, not significant. All experiments were repeated independently three times. Scale bars, 5 μ m.

BASL is also required, as demonstrated by the abolition of POLAR polarization with BASL loss-of-function⁴.

To provide genetic evidence for the function of POLAR as a BIN2 scaffold, we analysed the abaxial epidermis of cotyledons of *Arabidopsis* plants overexpressing POLAR in its endogenous domain and in plants overexpressing BIN2 in the stomatal lineage at 3 and 21 d.p.g. (Fig. 3, Extended Data Figs. 8, 9). At 3 d.p.g., similar to the *basl-2* mutant, there was an excess of small stomatal lineage cells in both genotypes. Treatment with the GSK3 kinase inhibitor bikinin (BIK)¹⁸ suppressed the phenotypes caused by overexpressing BIN2 (Fig. 3a, c, Extended Data Fig. 8c, d). However, whereas asymmetry was strongly reduced in the small epidermal dividing cells of the *basl-2* mutant, the same cells both in POLAR- and in BIN2-overexpressing plants retained a high degree of asymmetry (Fig. 3b). Unlike POLAR-overexpressing plants, which exhibited an increased pavement cell density and a reduced stomatal index as in the *basl-2* mutant, BIN2 overexpression reduced pavement cell density and increased the number of clustered stomata (Fig. 3a, c, Extended Data Fig. 9), suggesting a role for BIN2 in promoting differentiation of both pavement cells and stomata²⁴. Moreover, combined POLAR and BIN2 overexpression led to a further increase in cell divisions in the stomatal lineage, resulting in higher pavement and stomatal cell densities (Extended Data Figs. 8a, 9). By contrast, the abaxial epidermis of *polar pl1* double mutant plants (Extended Data Fig. 7) displayed a strong reduction in small cell and pavement cell densities (Fig. 3a, c, Extended Data Figs. 5g, h, 9). In a *polar-2 pl1-1 pl2-1* triple mutant (where PL2 is encoded by *At3g09730*) (Extended Data Fig. 7), these phenotypes were further enhanced, indicating functional

redundancy in the POLAR family. The *bin2-3 atsk12* double mutant (Extended Data Fig. 4) exhibited both a reduced number of small stomatal lineage cells, which subsequently resulted in a reduced number of stomata, and an increased number of symmetrically dividing pavement cells (Fig. 3a, c, Extended Data Fig. 9). The quadruple *bin2-3 bil1 bil2 atsk13*-RNAi mutant²⁵ (BIL1 and BIL2 are encoded by *At2g30980* and *At1g06390*, respectively) severely decreased the number of small stomatal lineage cells and pavement cell and stomatal densities (Fig. 3a, c, Extended Data Fig. 9), similar to the phenotype of the *polar-2 pl1-1 pl2-1* triple mutant. Together, these phenotypes indicate that BIN2, and probably other ATSKs are required first at the cortical polarity site, to restrict MAPK signalling and promote ACD, and second in the nucleus of SLGCs or meristemoids, to limit cell division and to promote differentiation into pavement or guard cells, respectively. Similarly to BIN2, ATSK12 was able to phosphorylate YDA and SPCH in vitro (Extended Data Fig. 4c), suggesting that it has a redundant function in inhibition of the MAPK module during ACD. However, the importance of the different ATSKs for differentiation and for ACD may differ, as the nuclear localization of BIN2 in the abaxial cotyledon epidermis was more pronounced than that of ATSK12.

To investigate whether the kinase activity of BIN2 is required for POLAR function, we examined the localization of POLAR-GFP in plants grown on BIK (Fig. 4a). BIK treatment stabilized POLAR protein and enhanced its polarization (Fig. 4a–c). To examine whether BIN2 phosphorylates POLAR, we performed in vitro kinase assays and identified 16 residues that were phosphorylated by BIN2, of which one was observed in vivo (Fig. 4d, e, Supplementary Table 1). Phosphorylation by BIN2 was almost completely abolished in POLAR(27A), a POLAR mutant in which 27 residues, identified sites and putative GSK3 phosphorylation motifs¹⁴, were substituted with alanine (Supplementary Table 1). POLAR phosphorylation by BIN2 was reduced when all eight putative GSK3 phosphorylation motifs in POLAR were mutated (POLAR(8GSKA)), or when 12 residues in a truncated POLAR were replaced with alanine (POLAR(26–166/12A)) (Fig. 4d, e, Supplementary Table 1). Similarly to BIN2, ATSK12 also phosphorylated POLAR in vitro (Fig. 4f). Because inhibition of BIN2 stabilized POLAR, we hypothesized that phosphorylation of POLAR by BIN2 increased its turnover. Consistent with this hypothesis, when expressed at endogenous levels in the *polar-2* mutant, POLAR(27A)-GFP was stabilized at the plasma membrane and led to increased cell division (Fig. 4h–j), similar to that observed in plants overexpressing POLAR-GFP (Fig. 3, Extended Data Fig. 8d).

BIN2 also phosphorylated BASL in vitro (Extended Data Fig. 10a, b, Supplementary Table 1). Although we did not identify BASL Ser72 as a phosphorylation site in our assay, this residue has a key role in BASL polarization in vivo^{12,13} and forms part of a GSK3 phosphorylation motif¹⁴; we therefore tested whether BIN2 can phosphorylate BASL at Ser72. The Ser72Ala mutation reduced BIN2 phosphorylation of BASL(1–84) (Extended Data Fig. 10a), indicating that BIN2 kinase activity, which is likely to operate redundantly with MAPK^{12,13}, is required for BASL function. Unexpectedly, BIK treatment markedly reduced BASL-GFP levels and led to a localization resembling that of hyperphosphorylated BASL¹³ (Extended Data Fig. 10c). Indeed, BIK-mediated ATSK inhibition activated MAPKs, enhanced degradation of SPCH and reduced *SPCH*, *POLAR* and *BASL* transcript levels (Extended Data Fig. 10c–e).

In summary, we propose that the precise accumulation, complex assembly and localization of BASL, POLAR, BIN2 and other ATSKs regulate cell fate in the stomatal lineage (Extended Data Fig. 10f). MMCs express high levels of POLAR, whereas BIN2 is undetectable. During MMC maturation, BIN2 expression initiates BASL polarization in a redundant fashion alongside MAPKs. The accumulation of BIN2 and POLAR at the cortical BASL polarity site attenuates MAPK signalling and relieves BIN2 inhibition of SPCH in the nucleus. SPCH therefore accumulates and sustains POLAR and BASL transcription, enabling ACD to proceed. As POLAR expression is retained in the daughter SLGC, the BIN2–POLAR–BASL polarity module is

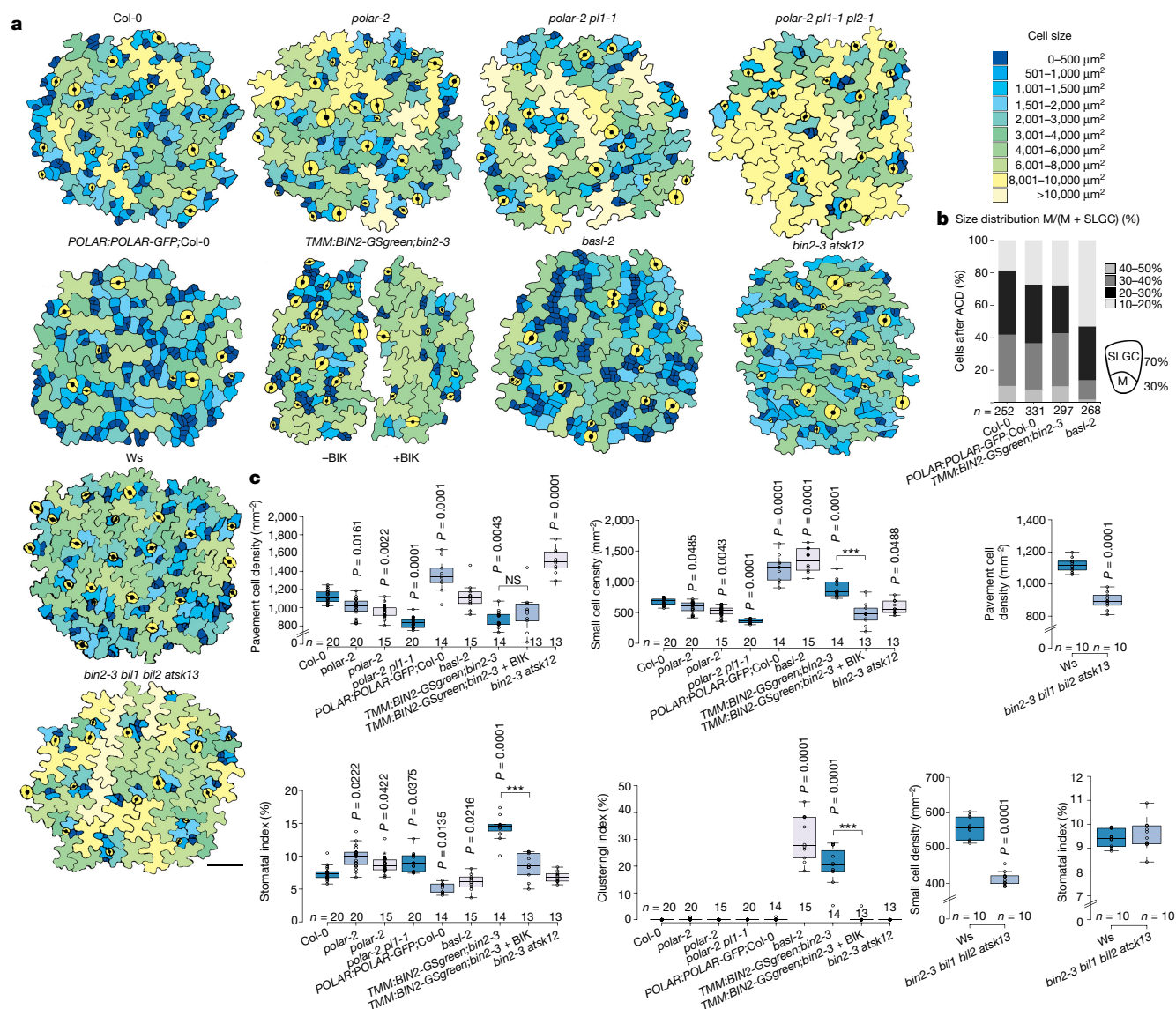


Fig. 3 | Stomatal phenotypes. **a**, Abaxial epidermis of cotyledons at 3 d.p.g. of wild-type, Col-0 and Wassilewskija (Ws) *Arabidopsis* plants, and other genotypes used in this study, as indicated. The TMM:BIN2-GSgreen;bin2-3 plants were grown on either 50 μM BIK or mock. Scale bar, 50 μm . Cell size distribution is presented as a colour scale. **b**, Quantification of the size distribution between meristemoids and SLGCs after cell division. *n*, number of cells from five biologically

independent cotyledons. **c**, Graphs representing the small cell (<1,000 μm^2) density, pavement cell density, stomatal index, and clustering for the genotypes in **a**. *n*, number of biologically independent cotyledons. One-way ANOVA with Tukey's post hoc test. All values were compared to Col-0 unless indicated with brackets. The *bin2-3 bil1 bil2 atsk13* quadruple mutant was compared to Ws. All experiments were repeated independently three times.

re-established at the opposite pole, enabling spacing ACD. Reduced POLAR expression in the SLGC leads to accumulation of BIN2 in the nucleus. This relieves the inhibition of MAPK signalling, resulting in SPCH degradation and pavement cell differentiation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0714-x>.

Received: 11 October 2017; Accepted: 25 September 2018;
Published online 14 November 2018.

- Pillitteri, L. J., Guo, X. & Dong, J. Asymmetric cell division in plants: mechanisms of symmetry breaking and cell fate determination. *Cell. Mol. Life Sci.* **73**, 4213–4229 (2016).
- Pillitteri, L. J. & Dong, J. Stomatal development in *Arabidopsis*. *Arabidopsis Book* **11**, e0162 (2013).
- MacAlister, C. A., Ohashi-Ito, K. & Bergmann, D. C. Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature* **445**, 537–540 (2007).

- Pillitteri, L. J., Peterson, K. M., Horst, R. J. & Torii, K. U. Molecular profiling of stomatal meristemoids reveals new component of asymmetric cell division and commonalities among stem cell populations in *Arabidopsis*. *Plant Cell* **23**, 3260–3275 (2011).
- Kim, T.-W., Michniewicz, M., Bergmann, D. C. & Wang, Z.-Y. Brassinosteroid regulates stomatal development by GSK3-mediated inhibition of a MAPK pathway. *Nature* **482**, 419–422 (2012).
- Khan, M. et al. Brassinosteroid-regulated GSK3/Shaggy-like kinases phosphorylate mitogen-activated protein (MAP) kinase kinases, which control stomata development in *Arabidopsis thaliana*. *J. Biol. Chem.* **288**, 7519–7527 (2013).
- Gudesblat, G. E. et al. SPEECHLESS integrates brassinosteroid and stomata signalling pathways. *Nat. Cell Biol.* **14**, 548–554 (2012).
- Lampard, G. R., MacAlister, C. A. & Bergmann, D. C. *Arabidopsis* stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. *Science* **322**, 1113–1116 (2008).
- Lampard, G. R., Lukowitz, W., Ellis, B. E. & Bergmann, D. C. Novel and expanded roles for MAPK signaling in *Arabidopsis* stomatal cell fate revealed by cell type-specific manipulations. *Plant Cell* **21**, 3506–3517 (2009).
- Lin, G. et al. A receptor-like protein acts as a specificity switch for the regulation of stomatal development. *Genes Dev.* **31**, 927–938 (2017).
- Dong, J., MacAlister, C. A. & Bergmann, D. C. BASL controls asymmetric cell division in *Arabidopsis*. *Cell* **137**, 1320–1330 (2009).

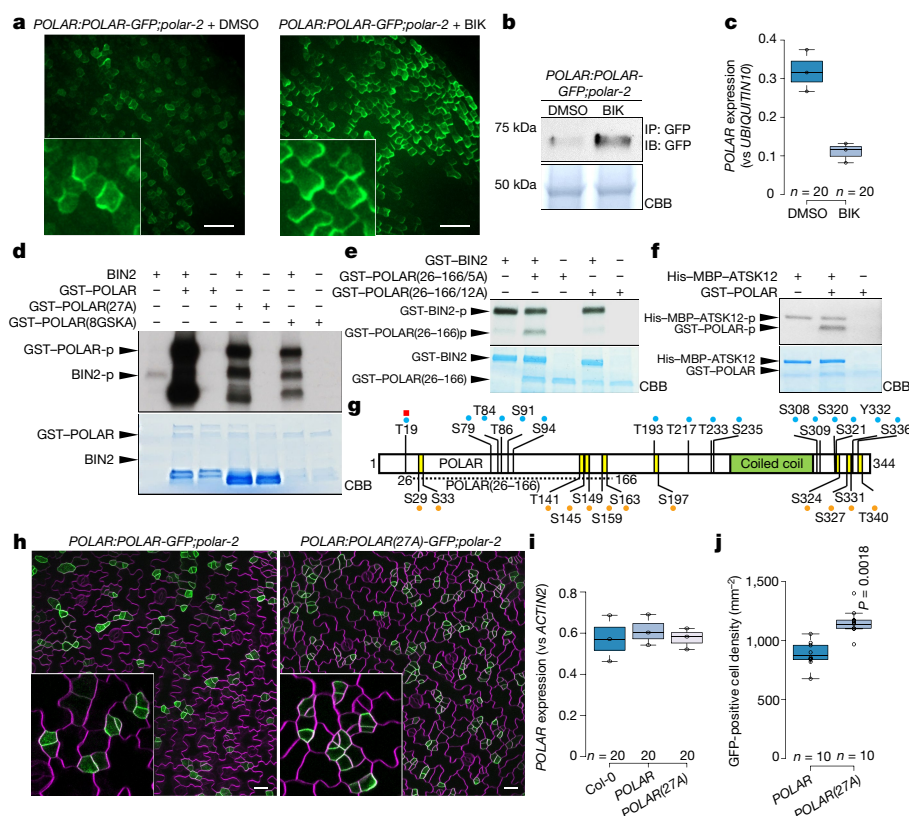


Fig. 4 | BIN2 activity is required for POLAR turnover. **a**, Confocal images of abaxial epidermis of cotyledons of POLAR-GFP plants grown on 50 μ M BIK or mock at 2.5 d.p.g. Scale bars, 50 μ m. **b**, Detection of POLAR-GFP protein by immunoprecipitation and immunoblot. **c**, POLAR gene expression of seedlings described in **a**. **d**, In vitro phosphorylation of POLAR by BIN2. POLAR(27A) and POLAR(8GSKA) exhibited reduced phosphorylation by BIN2. **e**, Truncated POLAR(26–166/12A) was not phosphorylated by BIN2. **f**, In vitro phosphorylation of POLAR by ATSK12. CBB, Coomassie brilliant blue. **g**, Schematic of the POLAR protein. Blue dots, BIN2 phosphorylation sites identified in

vitro; red square, BIN2 phosphorylation sites identified in vivo; orange dots, predicted BIN2 phosphorylation sites; yellow bands, putative GSK3 phosphorylation motifs; dashed line, POLAR(26–166). **h**, Confocal images of the abaxial epidermis of cotyledons of plants expressing POLAR-GFP and POLAR(27A)-GFP at 3 d.p.g. Scale bars, 20 μ m. **i**, POLAR gene expression of 20 seedlings in **h**. **c**, **i**, **n**, number of biologically independent seedlings. **j**, Cell density of GFP-positive cells in cotyledons shown in **h**. **n**, number of biologically independent cotyledons. One-way ANOVA with Tukey's post hoc test. All experiments were repeated independently three times. For blot source data, see Supplementary Fig. 1.

12. Zhang, Y., Wang, P., Shao, W., Zhu, J.-K. & Dong, J. The BASL polarity protein controls a MAPK signaling feedback loop in asymmetric cell division. *Dev. Cell* **33**, 136–149 (2015).
13. Zhang, Y., Guo, X. & Dong, J. Phosphorylation of the polarity protein BASL differentiates asymmetric cell fate through MAPKs and SPCH. *Curr. Biol.* **26**, 2957–2965 (2016).
14. Youn, J.-H. & Kim, T.-W. Functional insights of plant GSK3-like kinases: multi-taskers in diverse cellular signal transduction pathways. *Mol. Plant* **8**, 552–565 (2015).
15. Kim, T.-W. et al. Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat. Cell Biol.* **11**, 1254–1260 (2009).
16. Yan, Z., Zhao, J., Peng, P., Chihara, R. K. & Li, J. BIN2 functions redundantly with other *Arabidopsis* GSK3-like kinases to regulate brassinosteroid signaling. *Plant Physiol.* **150**, 710–721 (2009).
17. Rozhon, W., Mayerhofer, J., Petutsch, E., Fujioka, S. & Jonak, C. ASK1, a group-III *Arabidopsis* GSK3, functions in the brassinosteroid signalling pathway. *Plant J.* **62**, 215–223 (2009).
18. De Rybel, B. et al. Chemical inhibition of a subset of *Arabidopsis thaliana* GSK3-like kinases activates brassinosteroid signaling. *Chem. Biol.* **16**, 594–604 (2009).
19. Vert, G. & Chory, J. Downstream nuclear events in brassinosteroid signalling. *Nature* **441**, 96–100 (2006).
20. Blomme, J. et al. The mitochondrial DNA-associated protein SWIB5 influences mtDNA architecture and homologous recombination. *Plant Cell* **29**, 1137–1156 (2017).
21. Nadeau, J. A. & Sack, F. D. Control of stomatal distribution on the *Arabidopsis* leaf surface. *Science* **296**, 1697–1700 (2002).
22. Wakefield, J. G., Stephens, D. J. & Tavaré, J. M. A role for glycogen synthase kinase-3 in mitotic spindle dynamics and chromosome alignment. *J. Cell Sci.* **116**, 637–646 (2003).
23. Adrian, J. et al. Transcriptome dynamics of the stomatal lineage: birth, amplification, and termination of a self-renewing population. *Dev. Cell* **33**, 107–118 (2015).
24. Yang, K.-Z. et al. Phosphorylation of serine 186 of bHLH transcription factor SPEECHLESS promotes stomatal development in *Arabidopsis*. *Mol. Plant* **8**, 783–795 (2015).

25. Kondo, Y. et al. Plant GSK3 proteins regulate xylem cell differentiation downstream of TDIF-TDR signalling. *Nat. Commun.* **5**, 3504 (2014).

Acknowledgements We thank D. Bergmann, K. Torii, C. Grefen, Y. Kondo and H. Fukuda for materials; M. Pfeiffer, B. Pavie, J. Winkler, D. Van Damme, E. Mylle, S. Dhondt and A. Baekelandt for microscopy; M. De Cock for editing; V. Storme for statistics; B. De Rybel, J. Friml and N. Raikhel for critical reading. This work was supported by the Research Foundation—Flanders (G008416N) (E.R.), the China Scholarship Council (C.Z., K.W.), the Belgian Science Policy (K.W., M.T., M.K.Z., G.E.G.), the Agency for Innovation by Science and Technology (IWT) (M.K.Z.) and the Spanish Government Research Grant AGL2015-65053-R (M.M., C.F.).

Reviewer information Nature thanks S. Casson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author Contributions A.H., C.Z., M.T., K.W., A.d.M.S., D.V.S., M.J.U., M.K.Z., G.E.G., I.V., D.E., S.B., M.K. and C.B. performed experiments. T.J., C.F., M.M., S.d.V., G.D.J., A.H. and E.R. designed experiments. A.H. and E.R. wrote the manuscript and all authors revised it.

Competing Interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0714-x>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0714-x>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to E.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment. Experiments were repeated two to three times. All attempts at replication were successful.

Plant materials and growth conditions. *A. thaliana* L. (Heyhn.) (Columbia accession (Col-0)) seedlings and described genotypes were stratified for 2 days at 4 °C and germinated and grown on half-strength Murashige and Skoog (½MS) agar plates, supplemented with 1% (w/v) sucrose at 22 °C and a 16:8 h light:dark photoperiod for 3 or 21 days under 60 μmol m⁻² s⁻¹ of photosynthetically active radiation. When indicated, bikinin (Sigma-Aldrich) was added to the medium at a final concentration of 50 μM. The following mutant and transgenic *Arabidopsis* lines have been described previously: *basl-2*¹¹, *bin2-3* introgressed in Col-0⁷, *SPCH:SPCH-GFP;spch-3*⁷, *BASL:GFP-BASL;basl-2*¹¹, and *POLAR:POLAR-GFP;Col-0*⁴ and *bin2-3 bil1 bil2 atsk13-RNAi*²⁵ in the Wassilewskija (Ws) background. Transfer DNA (tDNA)-tagged mutants *polar*, *pl1-1*, *pl2-1* and *atsk12* were identified in the SALK and GABI-Kat tDNA collections (*polar-2*, SALK_146639; *pl1-1*, SALK_121087; *pl2-1*, SALK_123929 and *atsk12*, GK-410D09). Primers for genotyping and quantitative reverse transcription with PCR (qRT-PCR) are listed in Supplementary Table 2. The *polar-2 pl1-1*, *polar-2 pl1-1 pl2-1* and *bin2-3 atsk12* mutants were generated by crossing. Wild-type tobacco (*N. benthamiana*) plants were grown under 14 h of light and 10 h of darkness at 25 °C.

Preparation of constructs. The *POLAR*, *BASL*, *BIN2*, *TMM*, *ATSK11*, *ATSK12*, *ATSK13*, *BIL2* and *BIL1* promoters were isolated by PCR as fragments of 1,803 bp, 1,831 bp, 1,997 bp, 540 bp, 1,758 bp, 1,992 bp, 1,996 bp, 783 bp and 845 bp, respectively, upstream from the start codons, and were cloned into pDONR4-P1 (Invitrogen). For *POLAR* and *BIN2*, both cDNA and genomic DNA (gDNA) were cloned into pDONR221; for *BASL*, the cDNA was cloned into pDONR221 and the gDNA was cloned into pDONR221P3R (Invitrogen). For *ATSK11*, *ATSK12*, *ATSK13*, *BIL2*, *BIL1*, *ATSK31*, *ASK2*, *ATSK41* and *ATSK42*, the gDNA and cDNA were cloned into pDONR221. The kinase-dead *BIN2(K69R)* was obtained by site-directed mutagenesis of *BIN2* cDNA in pDONR221 with primers P21 and P22 (Supplementary Table 2), and the QuikChange II Site-Directed Mutagenesis Kit (Agilent) with TaKaRa DNA polymerase (Clontech). The gDNA of *BIN2*, *ATSK11*, *ATSK12*, *ATSK13*, *BIL2* and *BIL1* was recombined with the respective promoter in the pK7m34GW vector to generate the expression constructs with *BIN2:BIN2gDNA-GFP*, *ATSK11:ATSK11gDNA-GFP*, *ATSK12:ATSK12gDNA-GFP*, *ATSK13:ATSK13gDNA-GFP*, *BIL2:BIL2gDNA-GFP* and *BIL1:BIL1gDNA-GFP*, each carrying a kanamycin-resistance gene for plant selection. These constructs were introduced into the wild-type Col-0, and *BIN2:BIN2gDNA-GFP* was introduced into the single mutant *polar-2* and the double mutants *polar-2 pl1-1* and *polar-4 pl1-2*. *ATSK12:ATSK12gDNA-GFP* was introduced into the *atsk12* mutant. The cDNA of *BIN2* was recombined with the *TMM* promoter and the *GSgreen* tag²⁰ in the pK7m34GW and pH7m34GW vectors to generate two expression constructs, in *TMM:BIN2-GSgreen* containing kanamycin or hygromycin resistance, respectively. *TMM:BIN2-GSgreen* (kanamycin) was introduced into Col-0, whereas *TMM:BIN2-GSgreen* (hygromycin) was introduced into the *bin2-3* mutant.

POLAR gDNA was recombined with the *POLAR* promoter and the *GFP* in the pH7m34GW and with *mCherry* in the pB7m34GW vectors to generate the expression constructs *POLAR:POLAR-GFP* or *POLAR:POLAR-mCherry*, respectively. *POLAR* cDNA was recombined into the pK7FWG2 vector to generate 35S:*POLAR-GFP*. Both *POLAR:POLAR-GFP* and 35S:*POLAR-GFP* were introduced into Col-0, whereas *POLAR:POLAR-mCherry* was introduced into the transgenic plants expressing *TMM:BIN2-GSgreen*, *BIN2:BIN2gDNA-GFP*, *ATSK11:ATSK11gDNA-GFP*, *ATSK12:ATSK12gDNA-GFP*, *ATSK13:ATSK13gDNA-GFP*, *BIL2:BIL2gDNA-GFP* and *BIL1:BIL1gDNA-GFP* in Col-0. The phosphorylation mutant *POLAR(27A)* was generated by synthesis and was cloned into pDONR221. Both *POLAR* and *POLAR(27A)* were recombined with the *POLAR* promoter into pB7m34GW to generate the expression constructs *POLAR:POLAR-GFP* and *POLAR:POLAR(27A)-GFP*, which were used to complement the *polar-2* tDNA mutant. Both *BASL* gDNA and *SPCH* cDNA⁷ recombined with *BASL* and *SPCH*⁷ promoters in the pB7m34GW vector to generate the constructs *BASL:mCherry-BASL* and *SPCH:SPCH-mCherry*, respectively, which were transformed into plants expressing *TMM:BIN2-GSgreen* in Col-0. *BASL:mCherry-BASL* was also introduced into Col-0 plants expressing *BIN2:BIN2-GFP*. *BASL* gDNA was recombined with the *BASL* promoter in pH7m34GW to generate the construct *BASL:TagBFP2-BASL*, which was introduced into a line expressing *TMM:BIN2-GSgreen/POLAR:POLAR-mCherry* in Col-0.

For transient expression in tobacco, the wild-type and the kinase-dead *BIN2* cDNAs were recombined into the pK7FWG2 vector to generate the 35S:*BIN2-GFP* and 35S:*BIN2(K69R)-GFP* constructs. *ATSK11*, *ATSK12*, *ATSK13*, *BIL2*, *BIL1*, *ATSK3*, *ASK2*, *ATSK41* and *ATSK42* cDNA were recombined into the pK7FWG2 vector to generate 35S:*ATSK11-GFP*, 35S:*ATSK12-GFP*,

35S:*ATSK13-GFP*, 35S:*BIL2-GFP*, 35S:*BIL1-GFP*, 35S:*ATSK3-GFP*, 35S:*ASK2-GFP*, 35S:*ATSK41-GFP* and 35S:*ATSK42-GFP*, respectively. The *POLAR* cDNA was recombined with the 35S promoter and *mCherry* into the pB7m34GW vector to generate 35S:*POLAR-mCherry*. The *BASL* cDNA was recombined with the 35S promoter and *mCherry* into the pB7m34GW vector to generate 35S:*mCherry-BASL* and with the 35S promoter and *TagBFP2* to generate 35S:*TagBFP2-BASL*. The plasma membrane (UBQ10:Myr-TagBFP2-TagBFP2) and the nuclear (UBQ10:NLS-TagBFP2-TagBFP2) marker constructs were generated by recombining the *UBQ10* promoter containing either a myristoylation or a nuclear localization site sequence and pENL1-TagBFP2-L2 and pENR2-TagBFP2-L3 in the pB7m34GW vector. UBQ10:Myr-TagBFP2-TagBFP2 was introduced into Col-0 plants that expressed *TMM:BIN2-GSgreen*, *TMM:BIN2-GSgreen*; *POLAR:POLAR-mCherry* and *TMM:BIN2-GSgreen*; *SPCH:SPCH-mCherry*. For the rBiFC experiments²⁶, the cDNA of *BIN2*, *ATSK12*, *BASL* and *BKI1*²⁷ was cloned into pDONR221-P3P2, and cDNA of *POLAR*, *ATSK12*, *BASL* and *MPK6* was cloned into pDONR221-P1P4. pDONR221-P3P2-BIN2 and pDONR221-P3P2-ATSK12 were recombined with pDONR221-P1P4-POLAR and pDONR221-P3P2-BIN2 with pDONR221-P1P4-BASL into pBIFC-2in1-CN²⁶ to generate pBIFC-BIN2-nYFP+cYFP-POLAR, pBIFC-ATSK12-nYFP+cYFP-POLAR or pBIFC-BIN2-nYFP+cYFP-BASL, respectively. pDONR221-P3P2-BASL was recombined with pDONR221-P1P4-POLAR into pBIFC-2in1-NN or with pDONR221-P1P4-MPK6 or pDONR221-P1P4-ATSK12 into pBIFC-2in1-NC to generate pBIFC-nYFP-BASL+cYFP-POLAR, pBIFC-nYFP-BASL+MPK6-cYFP or pBIFC-nYFP-BASL+ATSK12-cYFP, respectively. pDONR221-P3P2-BKI1 was recombined with pDONR221-P1P4-POLAR into pBIFC-2in1-CN to generate pBIFC-BKI1-nYFP+cYFP-POLAR.

Construction of CRISPR vectors. To generate gRNA vectors, the entire chimeric RNA scaffold of the pEN-chimaera vector²⁸ was amplified with four primer pairs (Supplementary Table 2), flanked by *BsaI* recognition sequences and 4-bp Golden Gate-compatible overhangs²⁹. Amplified fragments were cloned into the corresponding sites of pGGA000, pGGB000, pGGC000 and pGGD000 to generate pGG-A-ATU6PTA-B and pGG-B-ATU6PTA-C, and pGG-C-ATU6PTA-D and pGG-D-ATU6PTA-E.

The PcUbi promoter and *Arabidopsis* codon-optimized Cas9 were PCR amplified from pDe-CAS9²⁸ and cloned into pGGB000 and pGGC000²⁹, generating pGG-A-PcUbi-B and pGG-C-Cas9PTA-D, respectively. The terminator G7T was amplified from pEN-R2-9-L3³⁰ and cloned into pGG000E to create pGG-E-G7T-F. The region containing the *ccdB* gene and chloramphenicol-resistant marker of pDONR221 was amplified with primers containing the attB (4 or 1)-*BsaI* site and overhang A and the attB (1 or 2)-*BsaI* site and overhang G, respectively, purified and used in a BP reaction (Invitrogen) with pDONR4P1R or pDONR221, to generate pEN-L4-AG-R1 and pEN-L1-AG-L2, respectively.

To generate linker plasmids, a cloning vector containing *SacI* and *ApaI* sites was made in a modified backbone of pDONR207 (Invitrogen) and designated pGGA1S1. *BsaI* sites in pGGA1S1 were removed by PCR mutagenesis. For each linker, a pair of sense and antisense oligonucleotides flanked by the *SacI* and *ApaI* sequences were annealed and cloned into the *SacI*-*ApaI* site of pGGA1S1 to generate pGG-A-Linker-C, pGG-C-Linker-G and pGG-E-Linker-G.

pEN-L4-AG-R1, pGG-A-PcUbi-B, pGGB003, pGG-C-Cas9PTA-D, pGGD002 and pGG-C-Linker-G were used in a Golden Gate assembly reaction to generate pEN-L4-PcUBI-Cas9PTA-G7T-R1. Of each plasmid, 100 ng was combined with 1× Cutsmart buffer (NEB), 1 mM ATP, 10 units of *BsaI* (NEB), and 400 units of T4 DNA ligase (NEB). The Golden Gate assembly reaction was carried out in a thermal cycler under the following conditions: 30 cycles of (37 °C for 3 min and 16 °C for 3 min); 50 °C for 5 min; 80 °C for 5 min.

Two unique gRNAs for *POLAR* (gRNA 1, TCGTGGGCCCGGTGGCTTTCCG; gRNA 2, TCGTGCAGTGTATACTCCT) and for *PLI* (gRNA 1, GTGTACAGATC TCTACCCCT; gRNA 2, ACATGTGTGTCCAGGACTAG) were selected based on the predicted mutation site in the first exon and minimum off-target scores using the CRISPR-P web tool³¹. Individual *POLAR* and *PLI* gRNAs were cloned into pGG-A-ATU6PTA-B and pGG-B-ATU6PTA-C and pGG-C-ATU6PTA-D and pGG-D-ATU6PTA-E, respectively. Of the *BbsI*-digested empty vector and of the annealed oligonucleotides, 50 ng and 1 μl were ligated with T4 DNA ligase (Invitrogen), according to the manufacturer's recommendations. Positive clones were identified by colony-touch PCR with the respective gRNA F and GGSeq_R primers. Plasmids were isolated with the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) and Sanger sequencing (VIB Genetic Service Facility) confirmed the sequence of the inserted gRNAs.

Golden Gate assembly was used to transfer *POLAR* and *PLI* gRNAs (1 and 2), and *POLAR* and *PLI* gRNA 1 into the recipient vector pEN-L1-AG-L2 in separate reactions. The gRNA Gateway entry clone was recombined with pEN-L4-PcUBI-Cas9PTA-G7T-R1 and the binary destination vector pK7m24GW³⁰ in a MultiSite Gateway reaction according to the manufacturer's recommendations.

Transformation and detection of Cas9-induced mutations in plants. *Arabidopsis* Col-0 plants were transformed with the *Agrobacterium tumefaciens* C58 strain using floral dip³². Transgenic T1 plants were selected on agar plates containing ½MS salts (1.5% (w/v) sucrose), supplemented with 50 mg l⁻¹ kanamycin and gDNA was extracted³³. Regions surrounding the *POLAR* and *PL1* target loci were PCR amplified with gene-specific primers with iProof High-Fidelity DNA Polymerase (Bio-Rad) under the following conditions: 98 °C for 3 min; 35 cycles of 98 °C for 10 s, 55 °C for 10 s; 35 cycles of 72 °C for 30 s, and a final extension with 72 °C for 5 min. PCR products were Sanger sequenced and compared with reference sequences at the expected target sites with the CLC main workbench (Qiagen Bioinformatics). For the selection of Cas9-free mutants, segregating T2 plants were initially screened for the absence of the tDNA with Cas9-specific primers via PCR. Cas9-negative, homozygous CRISPR mutants were identified based on the presence of a single mutant allele in the sequencing chromatogram. Primers used for vector preparation and PCRs are listed in the Supplementary Table 2.

Microscopy. For differential interference contrast (DIC) images, organs were excised manually and fixed in ethanol:acetic acid 9:1 (v/v) for 16 h, which was replaced with 90% (v/v) ethanol, and rehydrated with ethanol dilutions with increasing water content: 70%, 50%, 30%, and 10% ethanol, and pure distilled water as a final step. All these incubations were done at room temperature and for 1 h each. Finally, a chloral hydrate:glycerol:water solution (8:1:2, w/v/v) was used to clear the tissues. For image acquisition, samples were observed under a Nikon Eclipse 90i upright microscope with DIC optics and a DXM1200C camera or on an Olympus BX51 with DIC optics and a Nikon digital sight DS-SM camera.

The abaxial side of developing cotyledons of *Arabidopsis* seedlings at 3 d.p.g. or infiltrated *N. benthamiana* leaves were analysed with a FluoView1000 (Olympus) or Leica SP8 confocal microscope. Images were captured at 405 nm, 488 nm and 559 nm laser excitation and 425–460 nm, 500–530 nm and 570–670 nm long-pass emission for TagBFP2, eGFP and mRFP/mCherry/propidium iodide staining (1 mg ml⁻¹) (Sigma-Aldrich). For images with the Leica SP8, the gating technology was applied for autofluorescence removal. For images taken of plants that express the *UBQ10:Myr-TagBFP2-TagBFP2* plasma membrane marker, the blue channel was processed with ImageJ after acquisition to remove any cytosolic signal leaking from the plasma membrane marker and to improve visualization of fusion proteins in the GFP and RFP channels, which remained unmodified. For time-lapse imaging, cotyledons of *Arabidopsis* seedlings at 2 d.p.g. were dissected and mounted as previously described³⁴. Images were acquired at 10 or 15 min intervals and time-lapse movies were generated with ImageJ at a speed of 6 frames per second (f.p.s.).

Quantitative analysis of epidermal phenotypes. For cell analysis of developing cotyledons of *Arabidopsis* seedlings at 3 d.p.g., pictures were acquired by DIC microscopy. Images were printed and cell shapes were drawn by hand with a black marker. Cell drawings were scanned and contrast was adjusted to obtain clear images of the epidermis. Pavement cell density (PCD; number of pavement cells/mm²), small cell density (number of small cells/mm²), and stomatal index (SI; number of stomata/total number of epidermal cells × 100) were calculated by counting cell types in an area of 0.125 mm² with the cell counter plug-in in ImageJ. One representative drawing for each genotype was selected and every cell surface area was measured with ImageJ. Cell surface areas were subdivided into 10 subgroups to attribute a colour code, depending on the cell area. For the analysis of the asymmetry degree after ACDs, the surface area of the presumably freshly divided cell pairs corresponding to a meristemoid and a SLGC was measured with the quick selection tool in ImageJ. The ratio was obtained by dividing the meristemoid size by that of the SLGC plus that of the meristemoid. For cell analysis of mature cotyledons of *Arabidopsis* seedlings at 21 d.p.g., pictures were acquired by DIC microscopy. SI, stomatal density (number of stomata per mm²) and PCD were calculated by scoring two areas of 0.4 mm² located at both sides of the median axis of the cotyledon as described previously³⁵ with slight modifications. The grid and image counter plug-ins in ImageJ were used for counting the different cell types. Three representative images per genotype were printed and cell shapes were drawn by hand with a black marker as for the 3-day-old cotyledons. Each drawing was processed with a script³⁶ that automatically recognizes pavement cells and measures their surface areas. As for 3-day-old cotyledons, the cell surface areas were subdivided into 10 subgroups to attribute a colour code depending on the pavement cell area. The cell size distribution was then calculated from a minimum of 500 cells per genotype.

Transient expression in *N. benthamiana* and image analysis. *Agrobacterium* strain C58, carrying the constructs of interest and a p19-harboring strain in the abaxial side of tobacco leaves, was coinfiltrated as described previously³⁷. Multiple infiltrated leaves were observed by confocal microscopy 3–5 days after infiltration. For all quantifications of BIN2, POLAR and BASL localizations, *N. benthamiana* leaves were coinfiltrated with either plasma membrane or nuclear markers fused to TagBFP2. To calculate the ratio between the nuclear signal and the remainder of the cell, only cells with the nucleus and the plasma membrane in the same focal plane were considered. *z* stacks of three slices with an interval of 3 µm were imaged,

superimposed, and used for quantification. First, the cell was selected with the polygon selection tool in ImageJ. A script was provided by B. Pavie (VIB Imaging Core Facility, Leuven) to quantify the intensity ratios. The script recognizes the nucleus automatically and creates a mask for it, allowing unbiased quantifications of nuclear versus cytoplasmic plus plasma membrane signals in ImageJ. For all plasma membrane quantifications, intensity values corresponding to either high or low polarity regions of similar length (represented in Extended Data Fig. 3) were normalized to the plasma membrane marker. This normalization allowed processing of single-slice images, because relative intensity values are representative of protein amounts and do not originate from different focal planes inside the cells.

Generation and purification of bacterially produced proteins. Wild-type and mutant *POLAR* and *BASL* and wild-type *BIN2* were cloned into the pGEX6P1 vector⁷. *YDA* was cloned into pDEST15 and *ATSK12* was cloned into pDEST-HisMBP. Mutant versions of *POLAR* and *BASL* were first created in silico with the CLC workbench and then generated through DNA synthesis by GenScript. To generate *POLAR*(26-166/5A), *POLAR*(20A) was used as template for PCR with the P27 and P28 primers. *POLAR*(26-166/12A) was generated with the same template and P29–P30 and P31–P32 primers. Truncated wild-type or mutant versions of *BASL* were generated with the primers P38–P39 and P40–P41, respectively. The resulting plasmids were transformed into *Escherichia coli* BL21 Rosetta2 (DE3) cells. GST-tagged proteins were purified with glutathione-sepharose 4B beads (GE-Healthcare) and the GST tag was cleaved or not, depending on the experiments, from GST-BIN2 with PreScission Protease (GE-Healthcare). All primers are listed in Supplementary Table 2.

Immunoprecipitation-mass spectrometry (IP-MS) analysis. The protein extract was prepared by grinding 9 g (3 × 3 g for technical replicates) of seedlings at 3 d.p.g. expressing *TMM:BIN2-GSgreen* in *bin2-3* and *POLAR:POLAR-GFP*⁴ in Col-0 in liquid nitrogen, with extraction buffer (50 mM Tris-HCl, at pH 7.5, 150 mM NaCl, 1% (v/v) NP-40, and complete protease inhibitor (Roche Diagnostics)). Protein extracts were sonicated three times for 15 s on ice with at least a 15-s pause in between. The NP-40 concentration was diluted to 0.2% (v/v) in the protein extract by adding the extraction buffer without detergent. The extract was centrifuged at 18,000g at 4 °C for 30 min and 100 µl anti-GFP magnetic beads were added (Miltenyi Biotec) and incubated for 1 h on a rotating wheel at 4 °C. The beads were collected on a µMACS Separator (Miltenyi Biotec) and washed four times with 200 µl extraction buffer containing 0.2% (v/v) NP-40. The proteins were eluted from the beads by adding 50 µl of 50 mM NH₄HCO₃ preheated at 95 °C. Protein samples were prepared for mass spectrometry and run on a Proxeon II nLC - LTQ-Orbitrap XL³⁸ (Thermo Fisher Scientific) by means of Xcalibur 2.1 for BIN2 and analysed by MaxQuant and Perseus³⁸ or on a LTQ VelosOrbitrap mass spectrometer for POLAR (Thermo Fisher Scientific). The nano-high-performance liquid chromatography (HPLC) system used was an UltiMate 3000 Dual Gradient HPLC system (Dionex), equipped with a nanospray source (Proxeon), coupled to an LTQ VelosOrbitrap mass spectrometer (Thermo Fisher Scientific), operated in data-dependent mode with a full scan in the Orbitrap, followed by tandem mass spectrometry (MS/MS) scans of the 12 most abundant ions in the linear ion trap. MS/MS spectra were acquired in the multistage activation mode, with subsequent activation on fragment ions resulting from the neutral loss of –98, –49 or –32.6 m/z for phosphorylation site analysis. For peptide identification, all MS/MS spectra were searched by means of Mascot 2.2.04 (Matrix Science) against the *Arabidopsis* Information Resource protein sequence database (https://www.arabidopsis.org/servlets/Search?action=new_search&type=protein). The carbamidomethylation on cysteine and the oxidation on methionine were set as fixed and as variable modifications, respectively. Monoisotopic masses were searched within unrestricted protein masses for tryptic, chymotryptic and unspecific (subtilisin digest) peptides. Peptide and fragment mass tolerances were set to ± 5 p.p.m. and to ± 0.5 Da, respectively, whereas the maximum number of missed cleavages was set at 2. The result was filtered to 1% false discovery rate by means of the Percolator algorithm integrated into the Proteome Discoverer (1.3.0.339; Thermo Scientific).

Co-immunoprecipitation (co-IP) assay. Proteins were extracted from 3 g of *Arabidopsis* seedlings at 3 d.p.g., ground in liquid nitrogen in extraction buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.2% (v/v) Triton X-100, 10% (v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF), complete EDTA-free protease inhibitor cocktail (Roche)). The extracts were centrifuged at 18,000g at 4 °C for 30 min. A 1.5-ml volume of extracts was incubated with 50 µl magnetic RFP-binding or GFP-binding protein beads (RFP/GFP-Trap-M; Chromotek) at 4 °C for 1 h. After incubation, the beads were washed three times with 1 ml of extraction buffer in a magnetic rack to pellet the beads. The washed beads were mixed with 40 µl 2× sample loading buffer (LDS; Thermo Fisher) and boiled for 10 min at 70 °C. Samples were separated by 12% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and analysed with an anti-RFP antibody (mouse monoclonal (6G6) to RFP 1:2,500), followed by an anti-mouse-horseradish peroxidase (HRP) antibody (sheep 1:10,000; GE Healthcare) for RFP-tagged proteins,

and with an anti-GFP–HRP-conjugated antibody (monoclonal anti-mouse 1:5,000; Miltenyi Biotec) for GFP-tagged proteins.

Proteins were extracted from 1 g of infiltrated tobacco leaves ground in liquid nitrogen with extraction buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 10 mM EDTA, 1 mM sodium molybdate, 1 mM NaF, 10 mM dithiothreitol, 0.5% (w/v) polyvinylpyrrolidone, 1% (v/v) protease inhibitor cocktail (Sigma–Aldrich; 1 tablet per 10 ml extraction buffer), 1% (v/v) NP-40). The extracts were centrifuged at 18,000g at 4 °C for 30 min and the supernatant was passed through a 40-µm cell filter. The immunoprecipitation was carried out as described for *Arabidopsis*. The washed beads were mixed with 40 µl × 2 LDS sample buffer and boiled for 10 min at 70 °C. Samples were separated by 12% SDS–PAGE and analysed by anti-GFP–HRP-conjugated or with anti-HA (rat monoclonal 1:1,000; Sigma–Aldrich) antibodies, followed by an anti-rat (goat 1:10,000; GE Healthcare) antibody for detection of HA-tagged proteins.

In vitro kinase assay, mass spectrometry and phosphopeptide analysis. In vitro kinase assays with recombinant proteins were carried out with 5 µCi [γ -³²P]ATP in 30 µl kinase buffer (50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 10 mM MgCl₂, and 10 µM cold ATP) for 30 min at room temperature. The reaction was stopped by addition of 10 µl of 4 × LDS buffer. Proteins were resolved by 10% SDS–PAGE. After nonradioactive in vitro kinase assays, proteins were separated by SDS–PAGE, stained with Coomassie blue and in-gel digested with trypsin as described previously³⁹. The digested peptide mixtures were injected into a liquid chromatography–tandem mass spectrometry system using an Ultimate 3000 RSLCnano LC (Thermo Fisher Scientific) in-line connected to a Q Exactive mass spectrometer (Thermo Fisher Scientific), as described previously⁴⁰. To detect in vivo phosphorylated residues in POLAR, the IP–MS experiment was repeated without enrichment for phosphorylated residues by means of the Q Exactive⁴⁰. All phosphopeptides were also inspected manually. Accepted phosphopeptides, the related Mascot Ion Score, and the precursor mass deviation are listed in Supplementary Table 1.

MAPK activation assay. Col-0 seedlings were germinated and grown for 2.5 days on ½MS medium supplemented or not with 50 µM BIK. Proteins were extracted with 50 mM Tris, pH 7.5, 200 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) Tween 20, 1 mM PMSF, 1 mM dithiothreitol, 1 × phosphatase inhibitor mixture 2 (Sigma–Aldrich) and 1 × protease inhibitor mixture (Sigma–Aldrich). Proteins (30 µg) were resolved on 7.5% polyacrylamide gels and transferred onto a polyvinylidene fluoride membrane (Bio–Rad). Primary antibodies against MPK3 (1:2,500; Sigma–Aldrich) and MPK6 (1:8,000; Sigma–Aldrich) and against phospho-44/42 MAP kinase (1:2,500; Cell Signaling Technology) were used with HRP-conjugated anti-rabbit as a secondary antibody (1:8,000; GE Healthcare).

qRT–PCR. RNA was extracted from 100 mg seedlings at 2.5 to 5 d.p.g. with TRIzol (Invitrogen), followed by on-column purification with the RNeasy mini kit (Qiagen). cDNA was generated with the iScript cDNA synthesis kit (Bio Rad). POLAR, BASL, SPCH, BIN2, ATSK12, PL1, PL2, UBIQUITIN10, ACTIN2 and ACTIN1 genes were amplified from 1,000 ng total RNA with the primers listed in Supplementary Table 2. SYBR green I qPCR master mix (Roche) was used on a LightCycler 480 (Roche). All experiments were performed in triplicate.

Statistical analysis. All *P* values were calculated in R with one-way ANOVA and Tukey's post hoc honest significant difference test. In all figures, the values

were **P* < 0.05, ***P* < 0.005, ****P* < 0.0005. All measurements in Figs. 1–4 and Extended Data Figs. 1–10 are shown as box plots displaying the first and third quartiles, split by the median (line) and mean (cross), with whiskers extending to a maximum of 1.5 × interquartile range beyond the box. Outliers are represented as dots outside the whisker region.

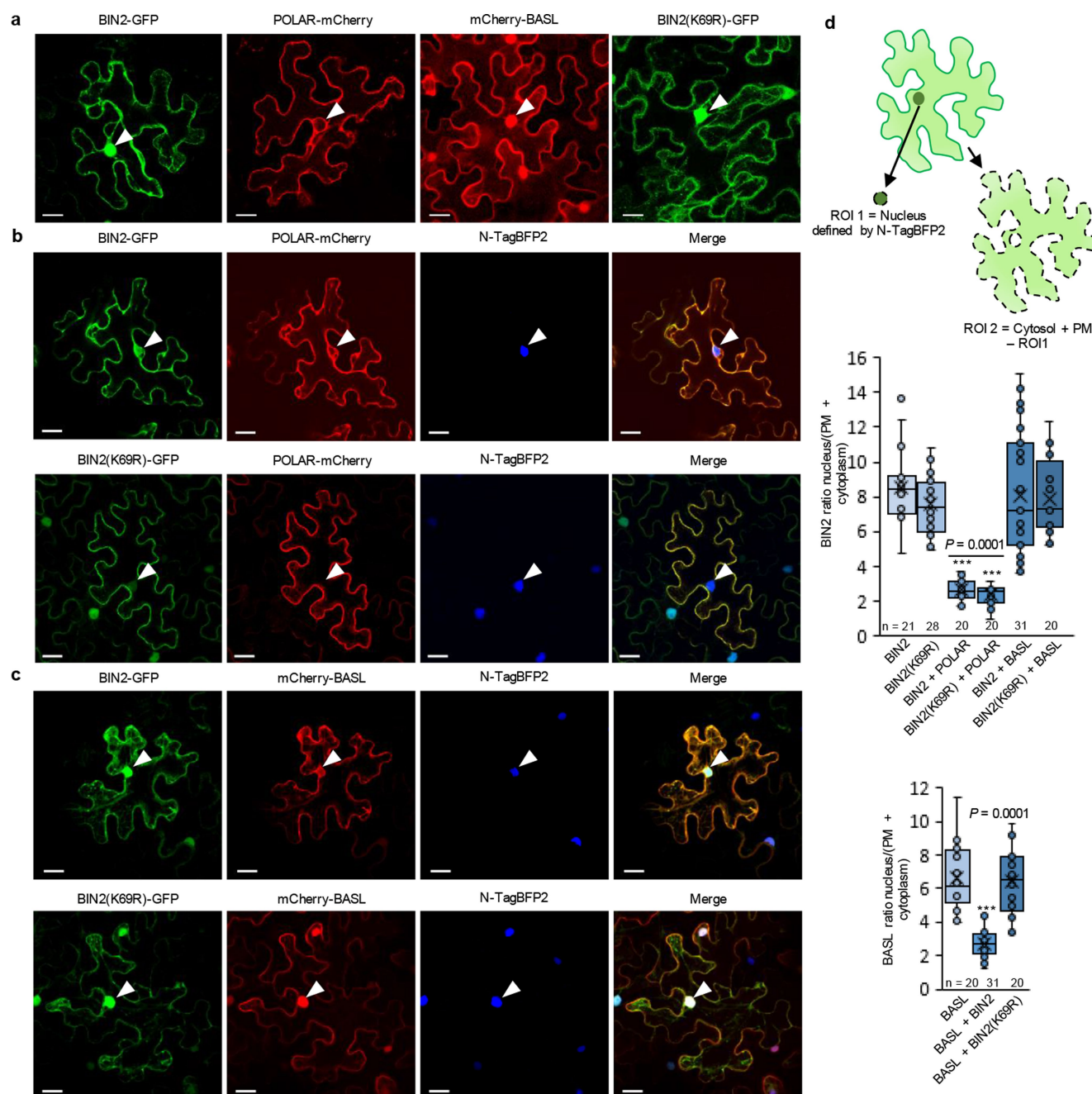
Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. Scripts used in this study are available upon request.

Data availability

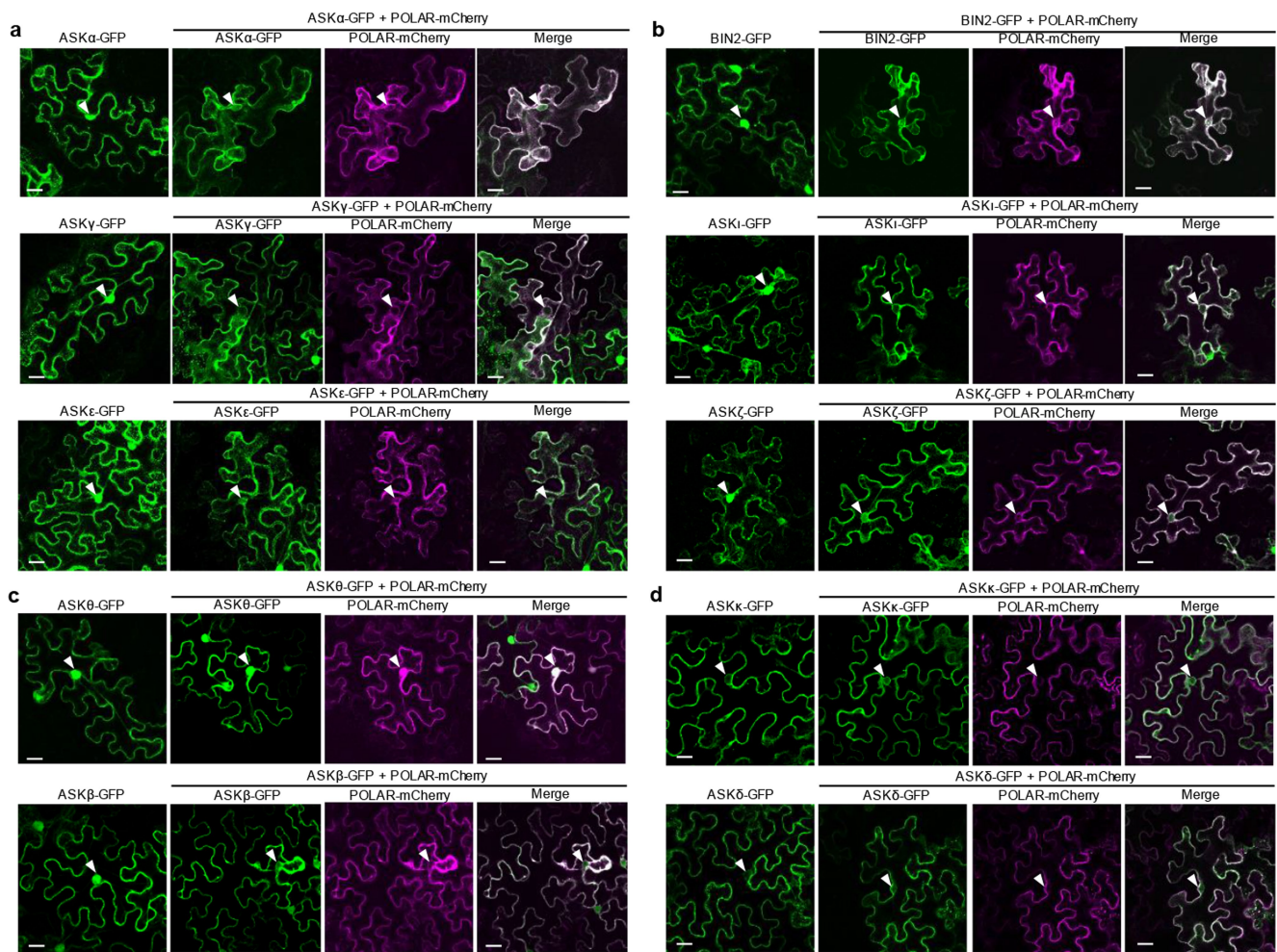
The datasets and accession codes supporting the findings of this study are available within the paper and its Supplementary Information. Source Data (gels and graphs) for Figs. 1–4 and Extended Data Figs. 1–10 are provided with the online version of the paper. There is no restriction on data availability.

26. Hecker, A. et al. Binary 2in1 vectors improve in planta (co)localization and dynamic protein interaction studies. *Plant Physiol.* **168**, 776–787 (2015).
27. Wang, X. & Chory, J. Brassinosteroids regulate dissociation of BK1, a negative regulator of BRI1 signaling, from the plasma membrane. *Science* **313**, 1118–1122 (2006).
28. Fauser, F., Schiml, S. & Puchta, H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* **79**, 348–359 (2014).
29. Lampropoulos, A. et al. GreenGate—a novel, versatile, and efficient cloning system for plant transgenesis. *PLoS One* **8**, e83043 (2013).
30. Karimi, M., De Meyer, B. & Hilson, P. Modular cloning in plant cells. *Trends Plant Sci.* **10**, 103–105 (2005).
31. Lei, Y. et al. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* **7**, 1494–1496 (2014).
32. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
33. Edwards, K., Johnstone, C. & Thompson, C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* **19**, 1349 (1991).
34. Peterson, K. M. & Torii, K. U. Long-term, high-resolution confocal time lapse imaging of *Arabidopsis* cotyledon epidermis during germination. *J. Vis. Exp.* **70**, e4426 (2012).
35. Delgado, D., Alonso-Blanco, C., Fenoll, C. & Mena, M. Natural variation in stomatal abundance of *Arabidopsis thaliana* includes cryptic diversity for different developmental processes. *Ann. Bot.* **107**, 1247–1258 (2011).
36. Andriankaja, M. et al. Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Dev. Cell* **22**, 64–78 (2012).
37. Boruc, J. et al. Functional modules in the *Arabidopsis* core cell cycle binary protein–protein interaction network. *Plant Cell* **22**, 1264–1280 (2010).
38. Wendrich, J. R., Boeren, S., Möller, B. K., Weijers, D. & De Rybel, B. In vivo identification of plant protein complexes using IP–MS/MS. *Methods Mol. Biol.* **1497**, 147–158 (2017).
39. Van Leene, J. et al. An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of *Arabidopsis* protein complexes. *Nat. Protoc.* **10**, 169–187 (2015).
40. Nelissen, H. et al. Dynamic changes in ANGUSTIFOLIA3 complex composition reveal a growth regulatory mechanism in the maize leaf. *Plant Cell* **27**, 1605–1619 (2015).



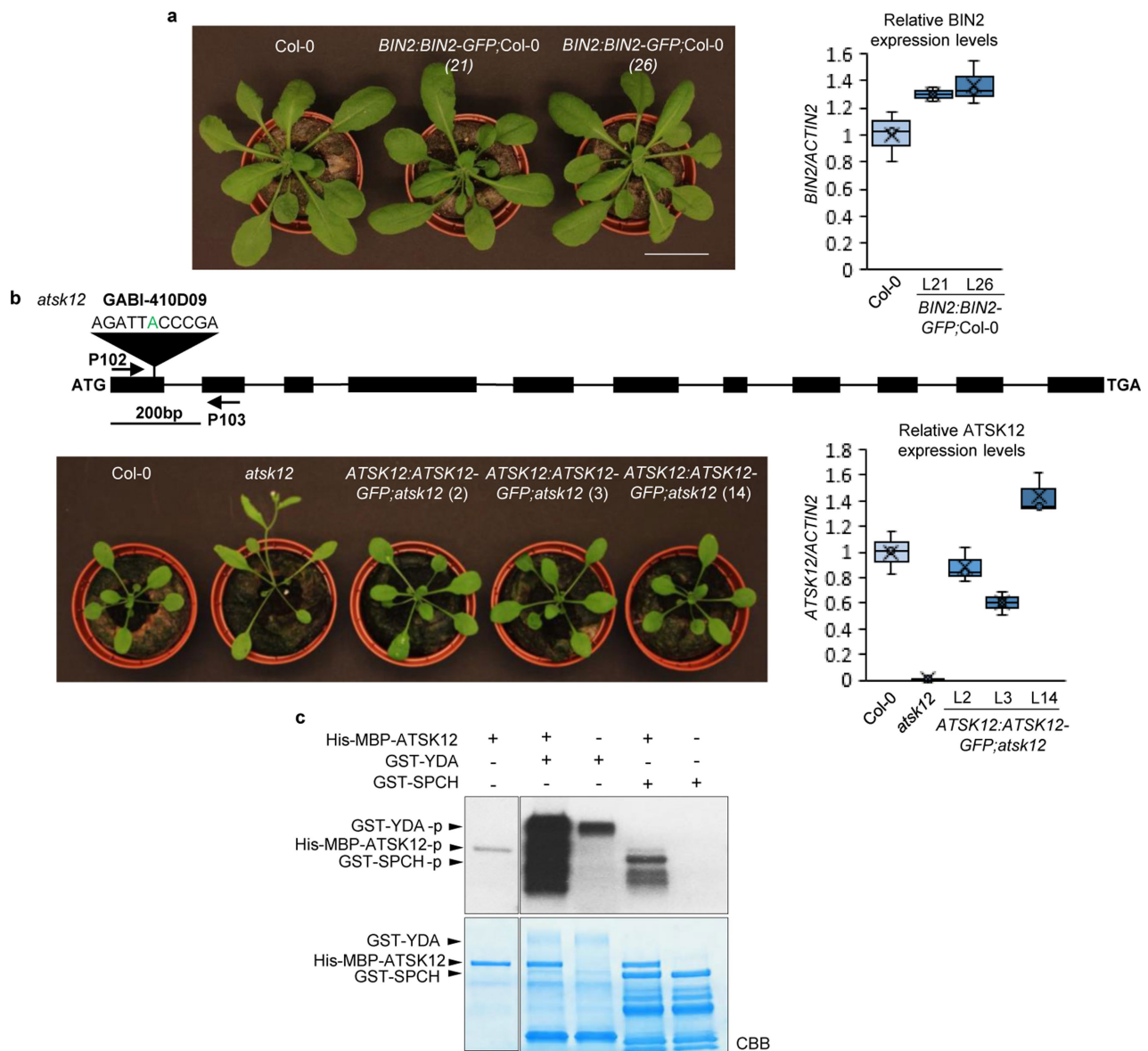
Extended Data Fig. 2 | Nuclear exclusion of BIN2 by POLAR and of BASL by BIN2. **a**, Transient expression in tobacco leaf epidermis of BIN2-GFP, POLAR-mCherry, mCherry-BASL, and the kinase-dead version of BIN2 (BIN2(K69R)-GFP). **b**, Transient co-expression of BIN2-GFP or BIN2(K69R)-GFP with POLAR-mCherry excluding BIN2 from the nucleus. **c**, Transient co-expression of BIN2-GFP with mCherry-BASL excluding BASL from the nucleus, whereas the transient co-expression of BIN2(K69R)-GFP with mCherry-BASL did not. The nucleus is indicated with arrowheads (**a**–**c**). **d**, Quantification of the nuclear exclusion of BIN2

and BASL. *n*, number of cells from three biologically independent leaves. Images were quantified with the nuclear (N)-TagBFP2 as a marker for the nucleus area and by means of a script to measure the average intensity in the nucleus compared to the remainder of the cell (see Methods). For **a**–**c**, experiments were repeated independently three times. Box plots show the first and third quartiles, split by the median (line) and mean (cross). Statistical analysis was performed with a one-way ANOVA and Tukey's post hoc test compared to BIN2-GFP or BIN2(K69R) or mCherry-BASL single infiltration, respectively. Scale bars, 20 μ m.



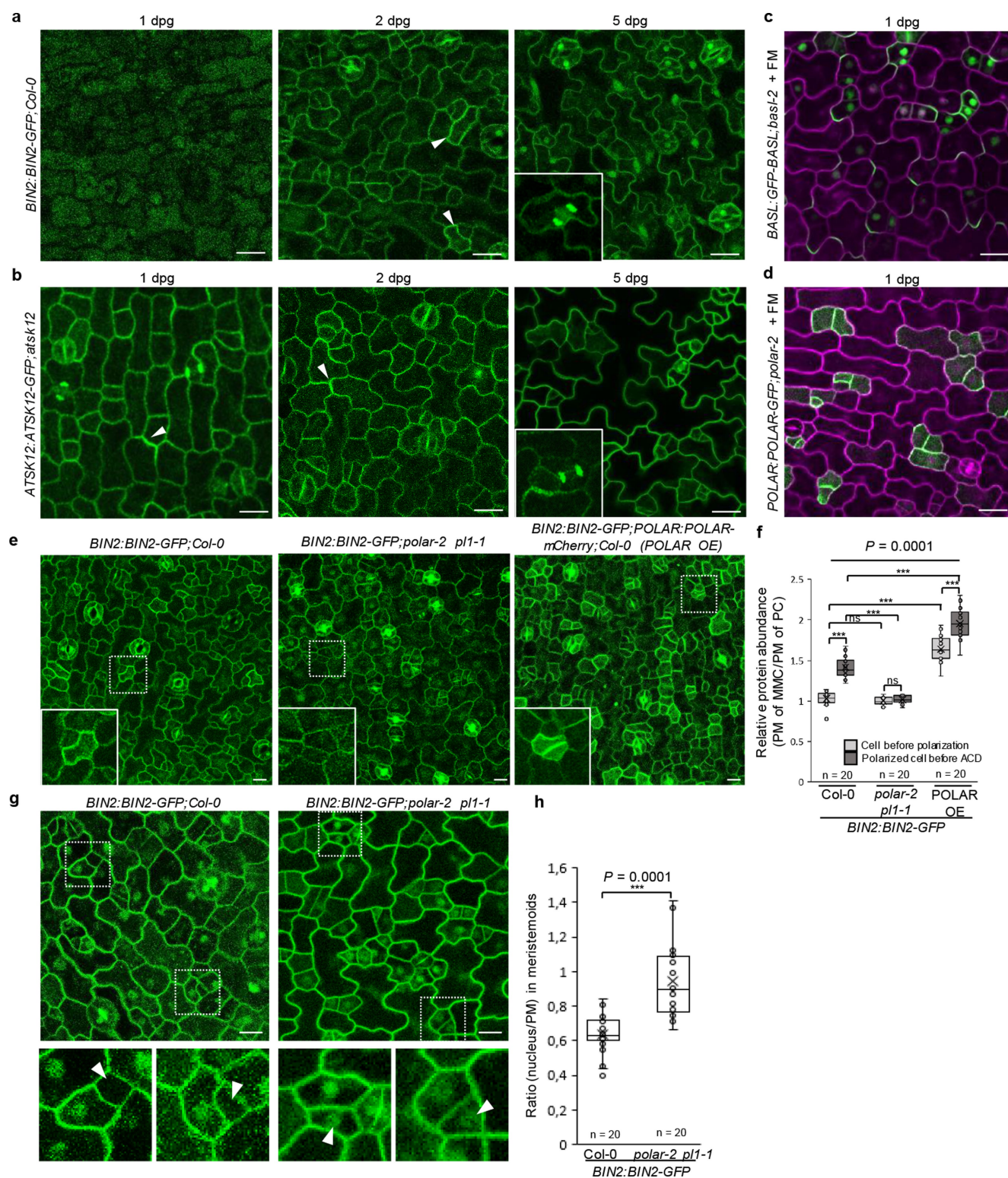
Extended Data Fig. 3 | Nuclear exclusion of other members of the *Arabidopsis* GSK3/SHAGGY family by POLAR. **a, b,** Transient co-expression in tobacco leaf epidermis of POLAR-mCherry with the group-I GSK3/SHAGGY-like kinases ATSK11(ASK α)-GFP, ATSK12(ASK γ)-GFP, and ATSK13(ASK ϵ)-GFP (**a**) and with the group-II GSK3/SHAGGY-like kinases BIN2(ASK η)-GFP, BIL2(ASK ι)-GFP, and (BIL1)ASK ζ -GFP (**b**) leading to their nuclear exclusion. **c,** Transient co-expression of POLAR-mCherry with the group-III GSK3/SHAGGY-like kinases ATSK32(ASK θ)-GFP and ATSK31(ASK β)-GFP. ATSK31(ASK β)-GFP was excluded from

the nucleus, whereas ATSK32(ASK θ)-GFP was not; POLAR-mCherry was included into the nucleus with ATSK32(ASK θ)-GFP. **d,** Transient co-expression of POLAR-mCherry with the group-IV GSK3/SHAGGY-like kinases ATSK41(ASK κ)-GFP and ATSK42(ASK δ)-GFP. When expressed alone, ATSK41(ASK κ)-GFP and ATSK42(ASK δ)-GFP did not localize to the nucleus, and coinfiltration with POLAR-mCherry did not affect their localization. White arrowheads in **a–d** indicate the nucleus. For **a–d**, experiments were repeated independently three times. Scale bars, 20 μ m.



Extended Data Fig. 4 | Characterization of *Arabidopsis* Col-0 and *atsk12* mutant endogenously expressing BIN2-GFP and ATSK12-GFP, respectively. **a**, Wild-type (Col-0) and *BIN2:BIN2-GFP;Col-0* plants (lines 21 and 26), grown for 4 weeks in soil with corresponding *BIN2* gene expression in seedlings measured by qRT-PCR. $n = 3$ biologically independent experiments. Scale bars, 2.5 cm **b**, Schematic representation of the tDNA insertions in the *atsk12* mutant (GABI-410D09). The insertion site is annotated with the flanking sequence and corresponding insertion site or inserted nucleotide in green. Primers used are depicted in the scheme and listed in Supplementary Table 2. Wild-type (Col-0) and

ATSK12:ATSK12-GFP;atsk12 complemented plants (lines 2, 3 and 14) grown for 2 weeks in soil with the corresponding *ATSK12* gene expression in seedlings measured by qRT-PCR. $n = 3$ biologically independent experiments. All qRT-PCR were carried out with seedlings at 5 d.p.g. Transcript levels were normalized to the *ACTIN2* gene expression. **c**, In vitro phosphorylation of YDA and SPCH by ATSK12. All experiments were repeated independently three times. Box plots show the first and third quartiles, split by the median (line) and mean (cross). For blot source data, see Supplementary Fig. 1.

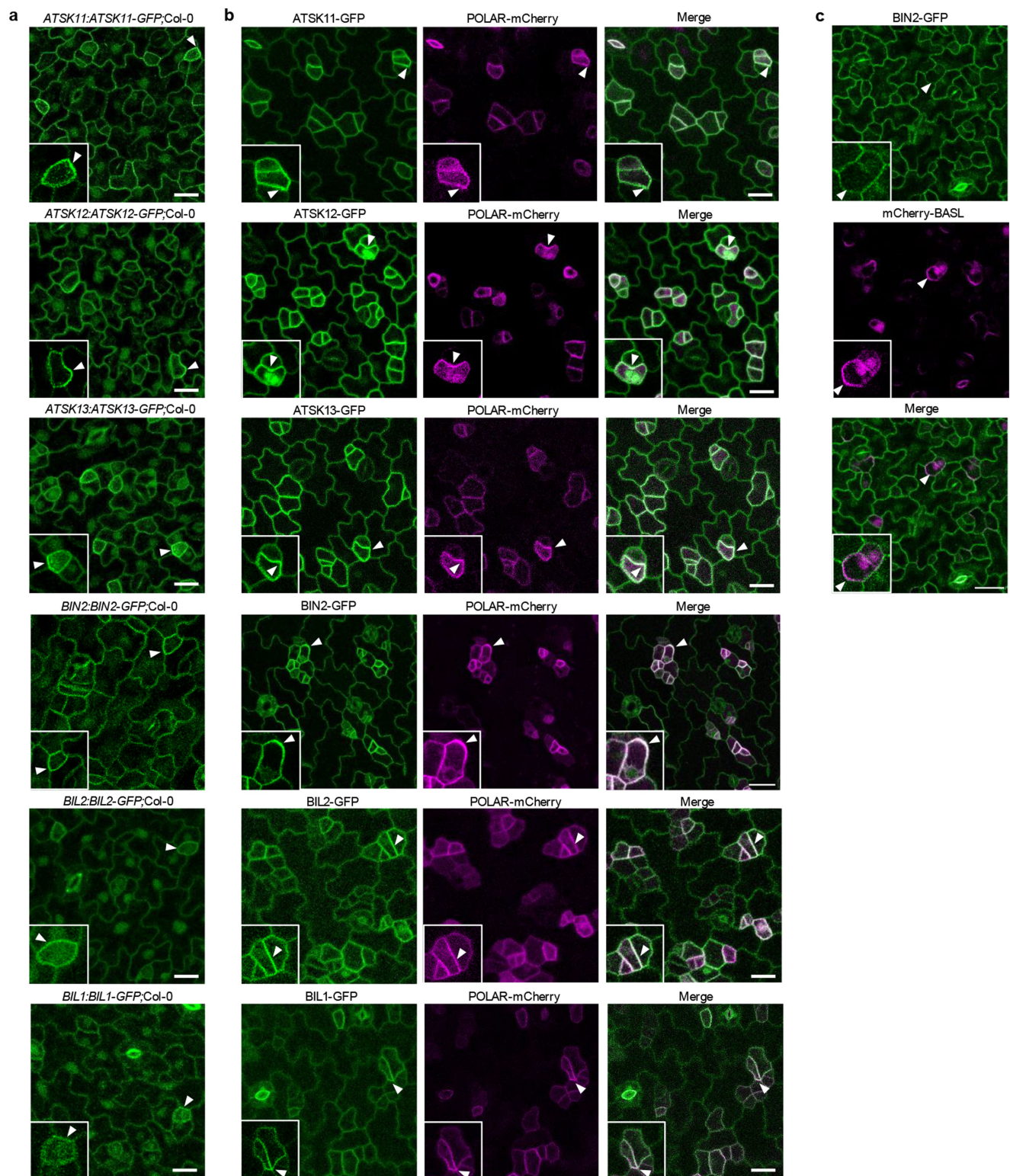


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Localization of BIN2, ATSK12, BASL and

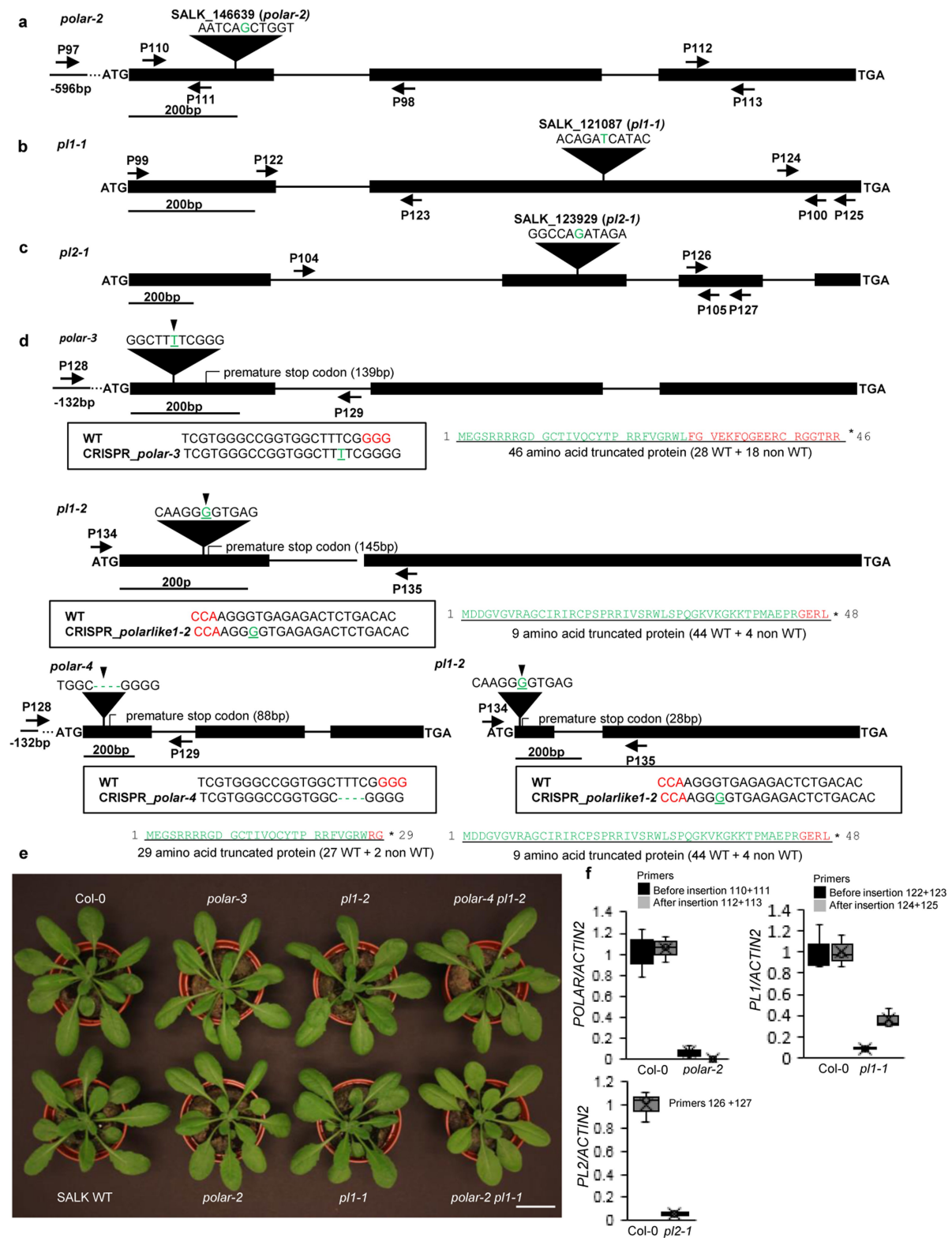
POLAR in *Arabidopsis*. **a**, Localization of BIN2 in the abaxial epidermis of *Arabidopsis* seedlings at 1, 2 or 5 d.p.g. BIN2 was not detectable at 1 d.p.g., but was visible at 1.5–2 d.p.g. BIN2 localized to the plasma membrane (PM), cytoplasm and nucleus. BIN2 was more abundant and polarized at the cell cortex before every ACD. **b**, Localization of ATSK12 in the abaxial epidermis of *Arabidopsis* seedlings at 1, 2 or 5 d.p.g. ATSK12 was detected at all developmental stages and localized to the plasma membrane and cytoplasm, but unlike BIN2, it had no nuclear signal. ATSK12 was also more abundant and polarized at the cell cortex before every ACD. Insets in **a** and **b** depict BIN2 and ATSK12 localization during cytokinesis. During mitosis, both proteins relocated to two distinct foci resembling the spindle localization as observed for mammalian GSK3. **c**, Localization of BASL in the abaxial epidermis of *Arabidopsis* seedlings at 1 d.p.g. **d**, Localization of POLAR in the abaxial epidermis of *Arabidopsis* seedlings at 1 d.p.g. **e**, BIN2 localization in POLAR loss- and gain-of-function mutants. BIN2 abundance in stomatal lineage cells undergoing ACD depends on POLAR. Loss of POLAR and PL1 markedly reduced, whereas POLAR overexpression greatly enhanced, BIN2 enrichment

before the ACD. Insets represent zoomed in regions. **f**, Quantification of the abundance of BIN2–GFP in the plasma membrane of cells undergoing ACD. *n*, number of cells from three biologically independent cotyledons. The intensity of BIN2–GFP was measured over time in the plasma membrane in stomatal lineage cells undergoing ACD. The plasma membrane signal of meristemoid cells was divided by the plasma membrane signal of neighbouring differentiated pavement cells to quantify the relative protein abundance. **g**, Nuclear localization of BIN2–GFP in Col-0 and the double tDNA *polar-2 pl1-1* mutant in the abaxial epidermis at 2 d.p.g. Meristemoids of the double *polar-2 pl1-1* mutant showed the increased nuclear localization of BIN2–GFP. The pictures below each main image represent zoomed in regions. White arrowheads indicate the nucleus in meristemoids. **h**, Quantification of the nuclear signal for BIN2 in **g**. *n*, number of cells from three biologically independent cotyledons. All experiments were repeated independently three times. Box plots show the first and third quartiles, split by the median (line) and mean (cross). One-way ANOVA with Tukey's post hoc test was used as indicated by brackets. Scale bars, 20 μ m.



Extended Data Fig. 6 | POLAR stabilized group I and group II ATSKs in the stomatal lineage. **a**, Confocal images of abaxial epidermis of cotyledons of seedlings at 3 d.p.g. expressing group-I *ATSK11:ATSK11-GFP;Col-0* (AT5G26751), *ATSK12:ATSK12-GFP;Col-0* (AT3G05840) or *ATSK13:ATSK13-GFP* (AT5G14640) and group-II *BIN2:BIN2-GFP;Col-0*, *BIL2:BIL2-GFP;Col-0* (AT1G06390) and *BIL1:BIL1-GFP;Col-0* (AT2G30980). *ATSK11-GFP*, *ATSK12-GFP*, *ATSK13-GFP*, *BIN2-GFP*, *BIL2-GFP* and *BIL1-GFP* proteins showed stabilization in the stomatal

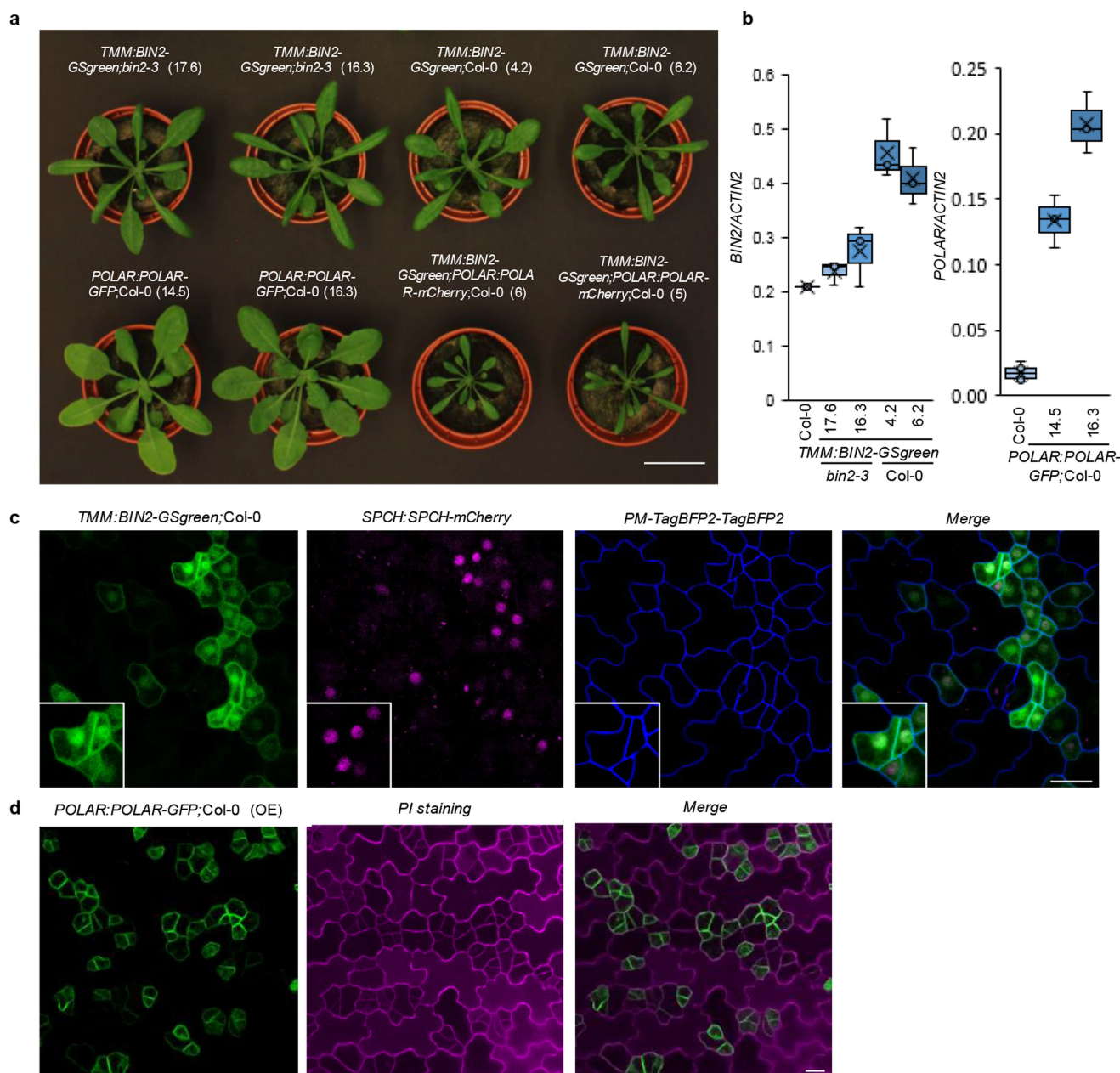
lineage and polarization during ACD, with group-I members displaying a more frequent polarization. **b**, Co-expression of all *BIN2* homologues with *POLAR-mCherry* leading to their stabilization and apparent polarization. White arrowheads in **a** and **b** indicate stabilization and/or polarization and correspond to insets in every picture. **c**, Co-expression of *BIN2* with *mCherry-BASL* did not lead to *BIN2* stabilization and polarization. All experiments were repeated independently three times. Scale bars, 20 μm .



Extended Data Fig. 7 | See next page for caption.

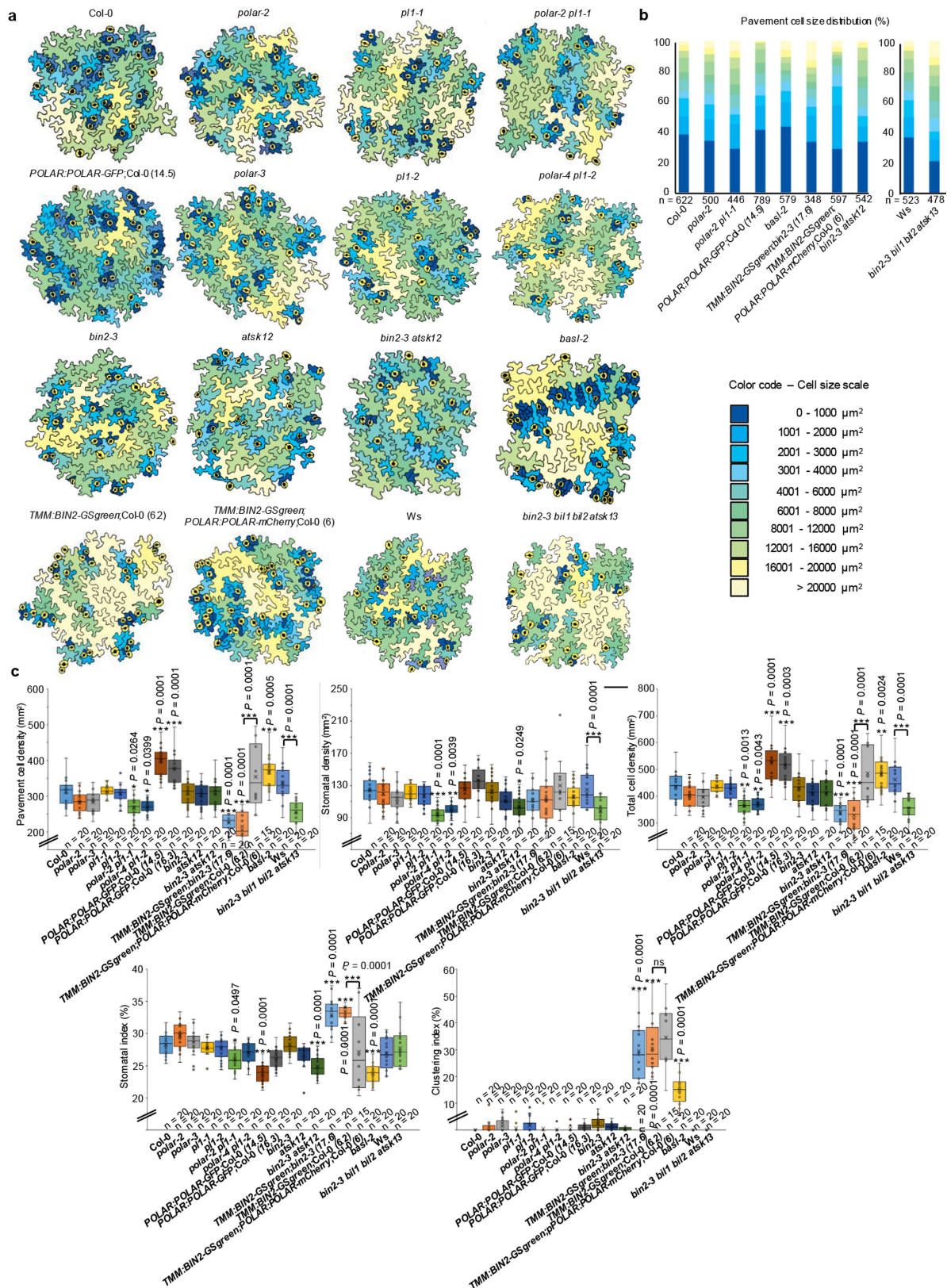
Extended Data Fig. 7 | Identification and characterization of POLAR, POLAR-like1 (PL1) and PL2 loss-of-function mutants. **a–c**, Schematic representation of the tDNA insertions in the *polar-2* (SALK_146639) (**a**), *pl1-1* (SALK_121087) (**b**) and *pl2-1* (SALK_123929) (**c**) mutants. **d**, Schematic representation of the POLAR and PL1 CRISPR mutants *polar-3* and *pl1-2*, respectively, and the double CRISPR mutant *polar-4 pl1-2*. Guide sequences of the selected target sites of the wild-type and of the CRISPR-edited *polar* are presented under the scheme. In the boxes, the protospacer-adjacent motif sequences are shown in red and inserted or deleted nucleotides in green. The resulting truncated amino acid sequence of the protein is shown below the scheme. Amino acids in green correspond to the wild-type POLAR protein and those in red to random

amino acids produced by the frame shift after the nucleotide insertion/deletion. **e**, Wild-type (Col-0), POLAR tDNA (*polar-2*) and POLAR CRISPR (*polar-3*) mutants and POLAR-like1 (PL1) tDNA (*pl1-1*) and PL1 CRISPR (*pl1-2*) mutants, as well as the corresponding double tDNA (*polar-2 pl1-1*) and CRISPR (*polar-4 pl1-2*) mutants grown for 4 weeks in soil. Scale bars, 2.5 cm. **f**, *POLAR*, *PL1* and *PL2* gene expression measured by qRT-PCR in seedlings of *polar-2*, *pl1-1* and *pl2-1* plants. $n = 3$ biologically independent experiments. Primers used are depicted in the scheme and listed in Supplementary Table 2. For **e** and **f**, experiments were repeated independently three times. Box plots show the first and third quartiles, split by the median (line) and mean (cross).



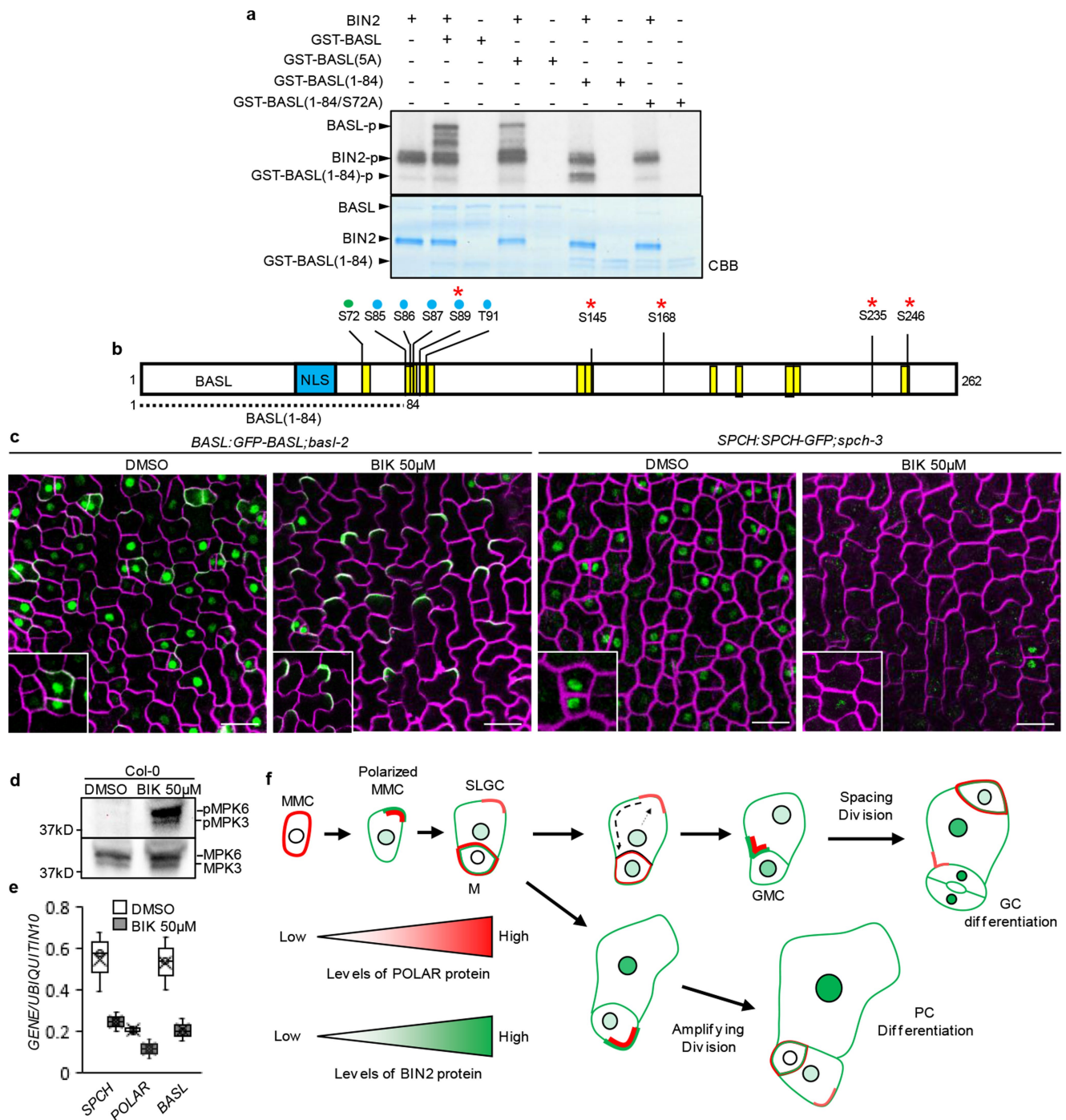
Extended Data Fig. 8 | BIN2 and POLAR overexpression in the stomatal lineage of *Arabidopsis*. **a**, *TMM:BIN2-GSgreen;bin2-3* (lines 17.6 and 16.3), *TMM:BIN2-GSgreen;Col-0* (lines 4.2 and 6.2), *POLAR:POLAR-GFP;Col-0* (lines 14.5 and 16.3) and *POLAR:POLAR-mCherry* (lines 6 and 5) introduced into *TMM:BIN2-GSgreen;Col-0* (line 6.2). Scale bar, 2.5 cm. **b**, *BIN2* and *POLAR* gene expression in the genotypes shown in **a**, measured by qRT-PCR. $n = 3$ biologically independent experiments.

c, SPCH-mCherry marking the stomatal lineage cell fate of the small dividing cells in *pTMM:BIN2-GSgreen;Col-0* (line 6.2) plants. **d**, Localization of POLAR-GFP in plants overexpressing *POLAR-GFP;Col-0* (line 14.5) presenting excessive divisions in the stomatal lineages. All experiments were repeated independently three times. Box plots show the first and third quartiles, split by the median (line) and mean (cross). Scale bars in **c** and **d**, 20 μ m.



Extended Data Fig. 9 | Stomatal phenotypes in abaxial epidermis of cotyledons at 21 d.p.g. a, Wild types (Col-0 or Wassilewskija (Ws)), *polar-2*, *pl1-1*, *polar-2 pl1-1*, *polar-3* (CRISPR), *pl1-2* (CRISPR), *polar-4 pl1-2* (CRISPR), *POLAR:POLAR-GFP;Col-0* (line 14.5), *bin2-3*, *atsk12*, *bin2-3 atsk12*, *bin2-3 bil1 bil2 atsk13-RNAi*, *TMM:BIN2-GSgreen;bin2-3* (line 17.6), *TMM:BIN2-GSgreen;Col-0* (line 6.2), *TMM:BIN2-GSgreen;POLAR:POLAR-mCherry;Col-0* (line 6), and *basl-2* seedlings. Scale bar, 100 μm . **b**, Pavement cell size distribution of genotypes described in **a** with significant phenotypes. *n*, number of cells from three biologically

independent cotyledons. **c**, Quantification of the pavement cell density, stomatal density, total cell density, the stomatal index, and the stomatal clustering index in the genotypes described in **a**. *n*, number of biologically independent cotyledons. Box plots show the first and third quartiles, split by the median (line) and mean (cross). Statistical analysis was done with a one-way ANOVA and Tukey's post hoc test. All values were compared to Col-0 unless indicated with brackets. The *bin2-3 bil1 bil2 atsk13-RNAi* quadruple mutant was compared to Ws. For **a**–**c**, experiments were repeated independently three times.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Phosphorylation of BASL by BIN2 and the working model for BIN2 and POLAR functions during ACD. **a**, In vitro phosphorylation of BASL by BIN2. Alanine substitutions in five BIN2 phosphorylated residues (BASL(5A)) reduced the BASL phosphorylation. The S72A substitution in the truncated BASL (BASL(1–84/S72A)) abolished phosphorylation. **b**, BASL protein. The blue dots, red asterisks, and green dot mark the identified in vitro BIN2 phosphorylation sites (this study), previously described MAPK phosphorylation sites, and the predicted S72 phosphorylation site, respectively. The nuclear localization signal (NLS) is marked in blue. The putative GSK3 phosphorylation motifs are marked in yellow. **c**, Confocal images of abaxial epidermis of cotyledons at 2.5 d.p.g. of *BASL:GFP-BASL;basl-2* or *SPCH:SPCH-GFP;spch-3* plants grown in the presence of 50 μ M bikinin (BIK) or mock (DMSO). Scale bars, 20 μ m. **d**, Wild-type (Col-0) seedlings at 2.5 d.p.g. grown in the presence of 50 μ M BIK or mock (DMSO). Western blot analysis of MPK3 and MPK6 activation upon treatment with BIK. Active MPK3 and MPK6 were detected by a phospho-p44/42 MAPK antibody. For blot source data, see Supplementary Fig. 1. **e**, Gene expression of *SPCH*, *POLAR* and *BASL* as in **d**, measured by qRT-PCR. $n = 3$ biologically independent experiments. Transcript levels were normalized

to the *UBIQUITIN10* gene. **f**, Model for the function of BIN2 and POLAR during ACDs. In the MMC or young meristemoid (M), POLAR is highly expressed, nonpolar, whereas the BIN2 expression is absent or undetectable. In mature MMCs or meristemoids, BIN2 is expressed, is upregulated and initiates BASL polarization redundantly with the MAPK module. The BIN2–POLAR–BASL complex migrates to the cell cortex and polarizes before the ACD, thereby relieving the BIN2 inhibition on SPCH in the nucleus and attenuating the MAPK signalling in the polar region. After the ACD, the smaller daughter cell (meristemoid) accumulates SPCH to sustain *POLAR* and *BASL* transcription, allowing further amplifying ACDs. When the *POLAR* expression remains in the larger daughter cell (SLGC), the BIN2–POLAR–BASL polarity module is reoriented at the opposite side, allowing the occurrence of spacing divisions. When the *POLAR* expression in the SLGC is low, BIN2 is more abundant in the nucleus, relieving its inhibitory function on the MAPK signalling, leading to SPCH degradation and pavement cell differentiation. Meristemoids destined for a guard mother cell fate lose the *POLAR* expression, while maintaining the *BIN2* expression. For **a–e**, experiments were repeated independently three times. Box plots show the first and third quartiles, split by the median (line) and mean (cross).

Sensitive tumour detection and classification using plasma cell-free DNA methylomes

Shu Yi Shen^{1,12}, Rajat Singhania^{1,12}, Gordon Fehrer^{2,12}, Ankur Chakravarthy^{1,12}, Michael H. A. Roehrl^{1,3,4}, Dianne Chadwick¹, Philip C. Zuzarte⁵, Ayelet Borgida², Ting Ting Wang^{1,4}, Tiantian Li¹, Olena Kis¹, Zhen Zhao¹, Anna Spreafico¹, Tiago da Silva Medina¹, Yadon Wang¹, David Roulois^{1,6}, Ilias Ettayebi^{1,4}, Zhuo Chen¹, Signy Chow¹, Tracy Murphy¹, Andrea Arruda¹, Grainne M. O'Kane¹, Jessica Liu⁴, Mark Mansour⁴, John D. McPherson⁷, Catherine O'Brien¹, Natasha Leigh¹, Philippe L. Bedard¹, Neil Fleshner¹, Geoffrey Liu^{1,4,8}, Mark D. Minden¹, Steven Gallinger^{9,10}, Anna Goldenberg¹¹, Trevor J. Pugh^{1,4}, Michael M. Hoffman^{1,4,11}, Scott V. Bratman^{1,4}, Rayjean J. Hung^{2,8*} & Daniel D. De Carvalho^{1,4*}

The use of liquid biopsies for cancer detection and management is rapidly gaining prominence¹. Current methods for the detection of circulating tumour DNA involve sequencing somatic mutations using cell-free DNA, but the sensitivity of these methods may be low among patients with early-stage cancer given the limited number of recurrent mutations^{2–5}. By contrast, large-scale epigenetic alterations—which are tissue- and cancer-type specific—are not similarly constrained⁶ and therefore potentially have greater ability to detect and classify cancers in patients with early-stage disease. Here we develop a sensitive, immunoprecipitation-based protocol to analyse the methylome of small quantities of circulating cell-free DNA, and demonstrate the ability to detect large-scale DNA methylation changes that are enriched for tumour-specific patterns. We also demonstrate robust performance in cancer detection and classification across an extensive collection of plasma samples from several tumour types. This work sets the stage to establish biomarkers for the minimally invasive detection, interception and classification of early-stage cancers based on plasma cell-free DNA methylation patterns.

The analysis of circulating tumour DNA (ctDNA) has numerous potential clinical applications. However, certain settings—such as cancer screening and the detection of minimal residual disease after treatment—require a degree of analytical sensitivity that is often beyond current technical limits of mutation-based ctDNA detection methods. The major obstacles to improved sensitivity of these methods include the limited number of recurrent mutations available to distinguish between tumour and normal circulating cell-free DNA (cfDNA) in a cost-effective manner, and technical artefacts (errors) introduced during sequencing. We reasoned that specific enrichment of methylated DNA fragments from cfDNA could overcome both of these issues.

To assess whether the higher number of DNA methylation changes in cancers could translate to increased sensitivity at lower sequencing costs, we performed bioinformatic simulations that examined the detection probability across varying numbers of differentially methylated regions (DMRs), coverage and ctDNA abundance (Fig. 1a, Extended Data Fig. 1a). We found improved sensitivity as the number of DMRs increased, even at lower sequencing depth and ctDNA abundance, which suggests that the recovery of cancer-specific DNA methylation changes could enable highly sensitive and low-cost detection, classification and monitoring of cancer.

However, this is challenging in practice owing to the low abundance and the fragmented nature of plasma cfDNA³, which have restricted

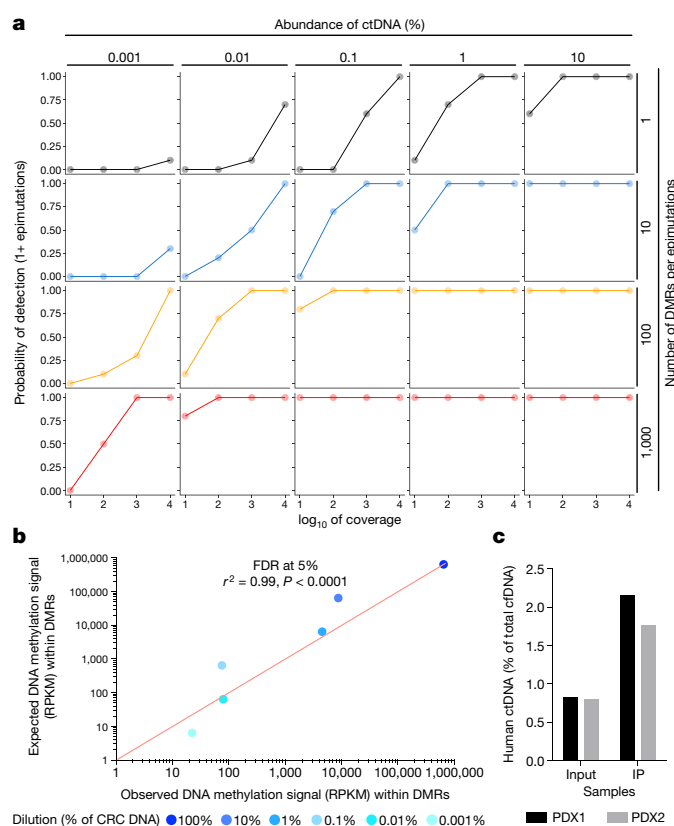


Fig. 1 | The cfDNA methylome as a sensitive approach to detect ctDNA in low levels of input DNA. a, Simulated probability of detecting at least one epimutation as a function of ctDNA concentration (0.001% to 10%; columns), number of DMRs analysed (1 to 10,000; rows) and sequencing depth (10× to 10,000×; x axis). **b**, Across a serial dilution series ($n = 7$ dilution points, two technical replicates, each replicate was used per protocol) of HCT116 DNA spiked into MM.1S multiple myeloma DNA, near-perfect correlations are observed between observed and expected methylation signal within DMRs in reads per kilobase of transcript per million mapped reads (RPKM). FDR at 5%, $r^2 = 0.99$; $P < 0.0001$. **c**, Frequency of ctDNA (human) as a percentage of total cfDNA (human + mouse) in the plasma from two colorectal cancer, patient-derived xenografts (PDX) before (input) and after (IP) cfMeDIP-seq.

¹Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. ²Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada. ³Memorial Sloan Kettering Cancer Center, New York, NY, USA. ⁴Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. ⁵Genome Technologies, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ⁶UMR_S 1236, Univ Rennes 1, Inserm, Etablissement Français du sang Bretagne, Rennes, France. ⁷Department of Biochemistry and Molecular Medicine, UC Davis Comprehensive Cancer Center, Sacramento, CA, USA. ⁸Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ⁹Fred Litwin Centre for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. ¹⁰Department of Surgery, Toronto General Hospital, Toronto, Ontario, Canada. ¹¹Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. ¹²These authors contributed equally: Shu Yi Shen, Rajat Singhania, Gordon Fehrer, Ankur Chakravarthy.

*e-mail: rayjean.hung@lunenfeld.ca; ddecary@uhnresearch.ca

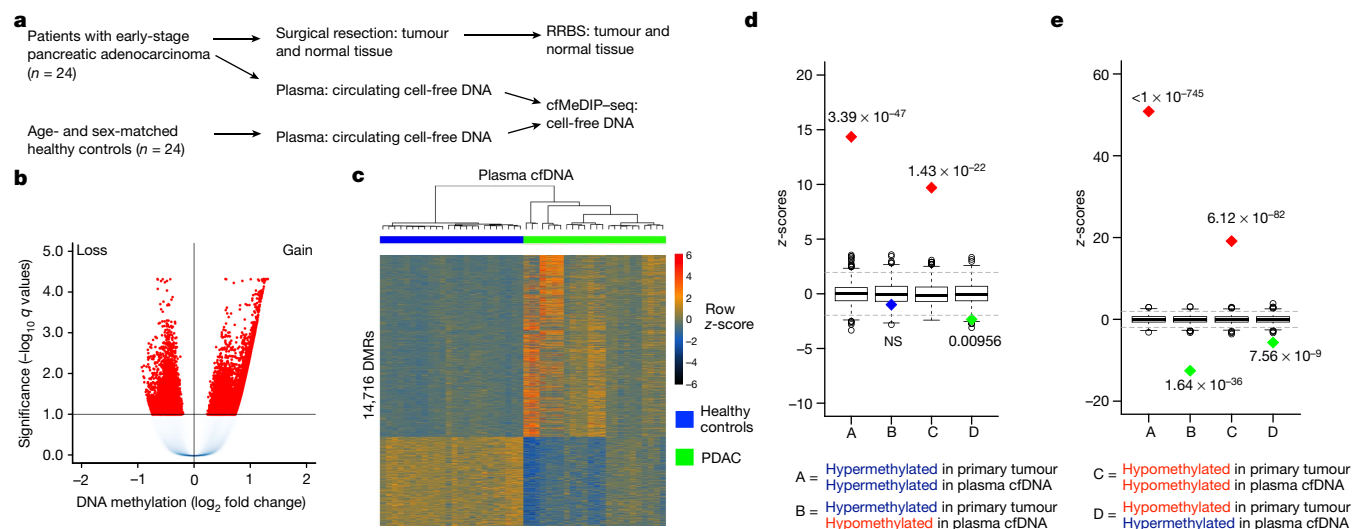


Fig. 2 | The cfMeDIP-seq method can identify thousands of DMRs in circulating cfDNA obtained from patients with pancreatic adenocarcinoma. **a**, Experimental design. **b**, Volcano plot of DMRs from patients with pancreatic cancer (cases, $n = 24$) versus healthy donors (controls, $n = 24$) using cfMeDIP-seq. Red dots indicate windows significant at BHFD < 0.1 (negative binomial GLM, two-sided P values). **c**, Heat map of the 14,716 DMRs identified in the plasma cfDNA from cases and controls (Euclidean distance, Ward clustering). Dendrogram shows separation by status (case or control). **d**, **e**, Overlap between case-

most of the previous plasma methylation profiling to locus-specific PCR-based assays^{7–9}. Although whole-genome bisulfite sequencing (WGBS) of cfDNA has been attempted^{10,11}, this approach is inefficient owing to degradation of around 84–96% of the input DNA during bisulfite conversion¹², high cost, and limited information recovery given the low genome-wide abundance of CpGs. Therefore, we developed cell-free methylated DNA immunoprecipitation and high-throughput sequencing (cfMeDIP-seq) for genome-wide bisulfite-free plasma DNA methylation profiling. This method can enrich CpG-rich, potentially more informative fragments, thus enhancing cost-effectiveness.

In brief, we optimized an existing low-input MeDIP-seq protocol¹³ that is robust down to 100 ng of input DNA, using exogenous *Enterobacteria phage* λ DNA (filler DNA) to increase the initial amount (Extended Data Fig. 1b). This is crucial for applications that are based on plasma cfDNA samples, which yield much less than 100 ng of cfDNA. We then performed extensive benchmarking of the optimized protocol. A comparison of low-input cfMeDIP-seq with gold-standard MeDIP-seq using colorectal cancer (CRC) HCT116 DNA that was sheared to mimic cfDNA showed robust CpG enrichment (Extended Data Fig. 2a–c) and inter-replicate correlation (Extended Data Fig. 2d). cfMeDIP-seq (1 to 10 ng input DNA) also recapitulated profiles from gold-standard MeDIP-seq (100 ng), reduced representation bisulfite sequencing (RRBS) (1,000 ng) and WGBS (2,000 ng) (Extended Data Fig. 2e).

Next, cfMeDIP-seq was compared to ultra-deep hybrid capture mutation sequencing based on unique molecular identifiers (UMIs)¹⁴ across a serial dilution of CRC DNA into multiple myeloma MM.1S cell-line DNA (Extended Data Fig. 3a). With cfMeDIP-seq, near-perfect linear associations were found between observed and expected numbers of DMRs (5% false discovery rate (FDR) threshold) and signals within DMRs, down to 0.001% dilution (both $r^2 = 0.99$, $P < 0.0001$) (Fig. 1b, Extended Data Fig. 3b–e). Hybrid capture mutation sequencing, however, detected CRC-specific mutations down to only 0.1% and 1% with single-strand consensus sequence (SSCS) and duplex consensus sequence (DCS), respectively (Extended Data Fig. 3f, g). This highlights the excellent analytical sensitivity of cfMeDIP-seq for the detection of cancer-derived DNA. We also evaluated the ability of cfMeDIP-seq to enrich ctDNA through biased sequencing of CpG-rich

versus-control plasma-derived DMRs and RRBS tumour-DMR-matched normal tissue (**d**) and PBMCs (**e**). Box plots represent the expected null distribution of overlaps from 1,000 permutations (two-sided, P values computed using standard normal distribution). The extremes of the boxes define the upper and lower quartiles and the centre lines define the median. Whiskers indicate $1.5 \times$ interquartile range (IQR). Diamonds represent observed overlap (red if significantly enriched, green if significantly depleted and blue if not significant). Horizontal lines indicate thresholds for statistical significance.

sequences that are frequently hypermethylated in cancer when compared to normal tissue¹⁵. Plasma from mice that carry patient-derived xenografts was used for cfMeDIP-seq, and a twofold enrichment of human-tumour-derived cfDNA was found after immunoprecipitation as compared to the input sample (Fig. 1c).

To investigate whether cfMeDIP-seq could detect ctDNA in early-stage cancer, we generated cfMeDIP-seq profiles from pre-surgery plasma cfDNA of 24 patients with primary early-stage pancreatic cancer (pancreatic ductal adenocarcinoma; PDAC) (cases) and 24 age- and sex-matched healthy controls (controls) (Fig. 2a, Extended Data Fig. 4a–f). In addition to plasma cfDNA, the microdissected primary tumours and adjacent normal tissue from the same patients with PDAC were used to generate DNA methylation profiles using RRBS. We identified 14,716 DMRs between the cfDNA of cases and controls (9,931 hypermethylated in cases, 4,785 in controls, based on negative-binomial generalized linear model (GLM) of fragment counts at a significance level of Benjamini–Hochberg FDR (BHFD) of 0.1) (Fig. 2b, c, Supplementary Table 1).

In comparison, 45,173 differentially methylated CpGs (DMCs) were found between tumour and normal tissue in RRBS data (Supplementary Table 2). Permutation testing to estimate the significance of overlaps between cfMeDIP-seq cell-free DMRs and RRBS tissue DMCs revealed significant enrichment for DMR and DMC pairs that are concordantly hypermethylated ($P = 3.39 \times 10^{-47}$) and concordantly hypomethylated ($P = 1.43 \times 10^{-22}$) in the case of cfDNA and tumour tissue. This significant enrichment was not observed in the discordant methylation pattern between cfDNA and tumour DNA (Fig. 2d). Furthermore, signals in overlapping plasma cfDNA and tissue DNA methylation were correlated (Extended Data Fig. 5a). These findings suggest that cfMeDIP-seq of plasma cfDNA can detect tumour-derived DNA methylation events in ctDNA.

As non-tumour-derived cfDNA is mostly released from blood cells, we performed similar permutation-based enrichment testing between case-versus-control cfMeDIP-seq DMRs and the 95,388 RRBS DMCs between PDAC tumour tissue ($n = 24$) and normal peripheral blood mononuclear cells (PBMCs) ($n = 5$) (Supplementary Table 3). Again, we observed significant enrichment for concordant hypermethylated ($P < 1 \times 10^{-745}$) and hypomethylated ($P = 6.12 \times 10^{-82}$) sites

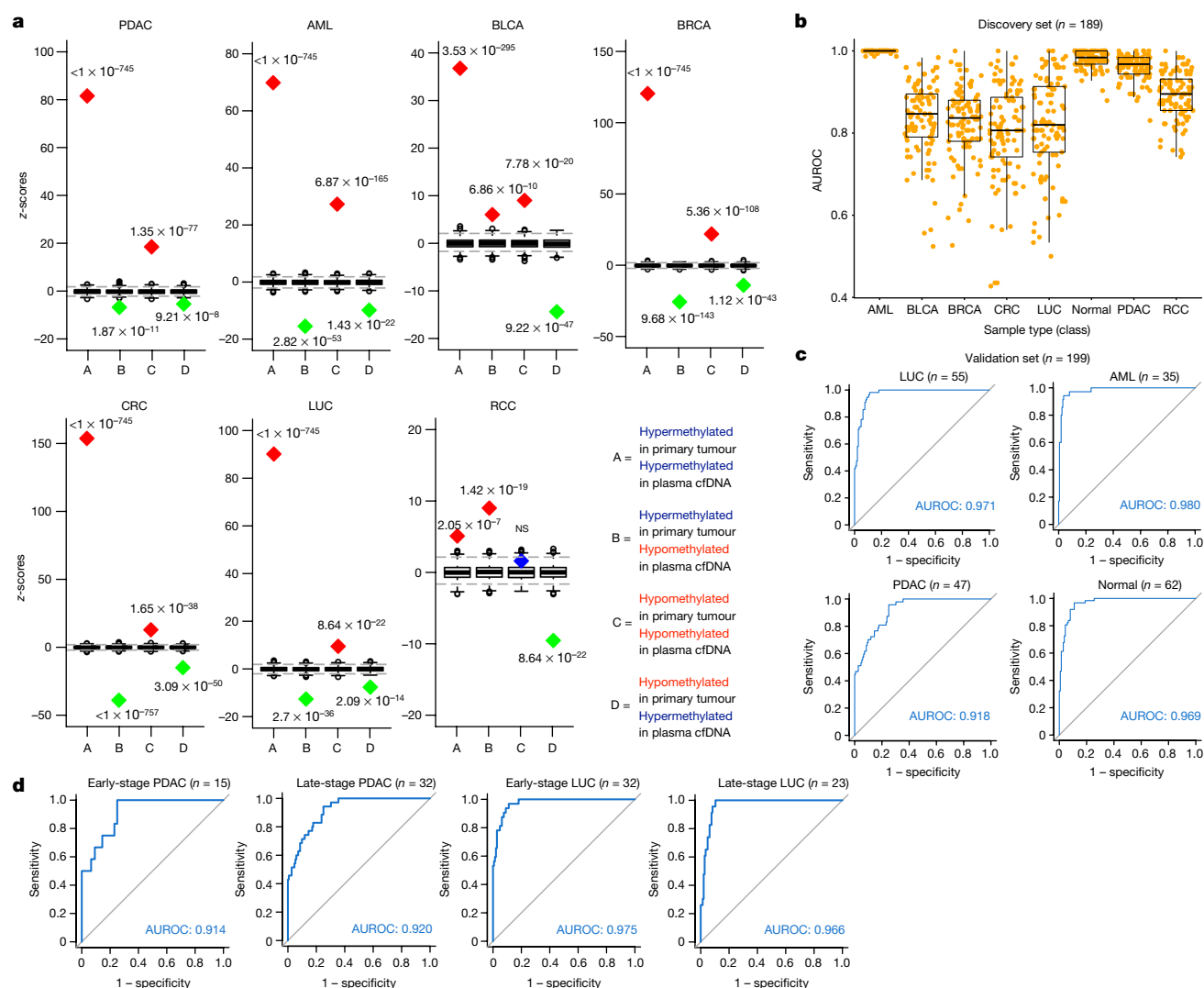


Fig. 3 | Methyome analysis of plasma cfDNA enables tumour classification. **a**, cfMeDIP-seq carried out on a discovery cohort consisting of 189 samples from seven different tumour types: PDAC, AML, BLCA, BRCA, CRC, LUC and RCC, including early- and late-stage tumours, and healthy controls (normal). For each cancer type, DMRs between the cancer type and normal controls were identified. Overlap is shown between plasma-derived DMRs for each cancer type and primary-tumour DMRs (tumour tissue versus adjacent normal tissue) for the corresponding cancer type using TCGA data. Box plots represent the expected null distribution of overlaps from 1,000 permutations (two-sided, P values computed using standard normal distribution). The extremes of boxes define the upper and lower quartiles and the centre lines define the medians. Whiskers indicate $1.5 \times$ IQR. Diamonds represent observed overlap (red if significantly enriched, green if significantly depleted and blue if not significant). Horizontal lines indicate thresholds for statistical significance. **b**, Evaluation of classification accuracy on the

discovery cohort. The discovery cohort ($n = 189$) was partitioned into 100 independent training and test sets in an 80%–20% manner, consisting of 8 classes (cancer types and healthy controls). Training sets were used for DMR selection and model training, yielding 100 sets of 8 one-class versus-other-classes binomial GLMnet classifiers. The y axis depicts distributions of AUROC for each held-out test set for each class. Dots represent performance in individual test sets. The extremes of boxes define the upper and lower quartiles and the centre lines define the medians. Whiskers indicate $1.5 \times$ IQR. **c**, ROC curves constructed using averaged class probabilities for independent validation set samples ($n = 199$, 55 LUC, 35 AML, 47 PDAC and 62 healthy controls) from the 100 models for each one-class-versus-other-classes comparison trained using the discovery cohort. **d**, ROC curves for the PDAC and LUC validation set divided into early and late stage, showing that the ability to discriminate PDAC or LUC samples is similar when considering early- and late-stage samples of that class separately.

in cfMeDIP-seq DMRs and tumour compared with PBMC DMCs, whereas discordant calls were underrepresented (Fig. 2e). In addition, signals in overlapping DMRs and DMCs were correlated (Extended Data Fig. 5b), and altogether indicated that DMRs identified using cfMeDIP-seq, between cases and controls, were probably derived from ctDNA (Extended Data Fig. 5c).

On the basis of the enrichment of tumour-derived DMRs and the known methylation-specific variable binding of transcription factors¹⁶, we hypothesized that cfMeDIP-seq methylomes could identify active transcriptional networks in tumours or other tissues using plasma cfDNA. Upon motif enrichment analysis on cfMeDIP-seq DMRs and taking methylation preferences of candidate transcription factors into

account¹⁶, we identified 42 transcription factors as binding in healthy controls and 52 as binding in cases of pancreatic cancer (Supplementary Tables 4, 5). As expected, the former included haematopoietic-lineage-specific transcription factors such as PU.1, NFE2 and GATA1, whereas the latter included the pancreas-associated transcription factors PTF1a, Onecut1 (HNF6) and NR5A2 (Extended Data Fig. 6a, c). Compared to random sets of transcription factors, those inferred as active in healthy controls are overexpressed in blood according to data from the Genotype-Tissue Expression (GTEx) project, whereas those inferred as active in cases of pancreatic cancer were found to be overexpressed in pancreatic tissues (according to GTEx data) and PDAC tissue (according to data from The Cancer Genome Atlas (TCGA);

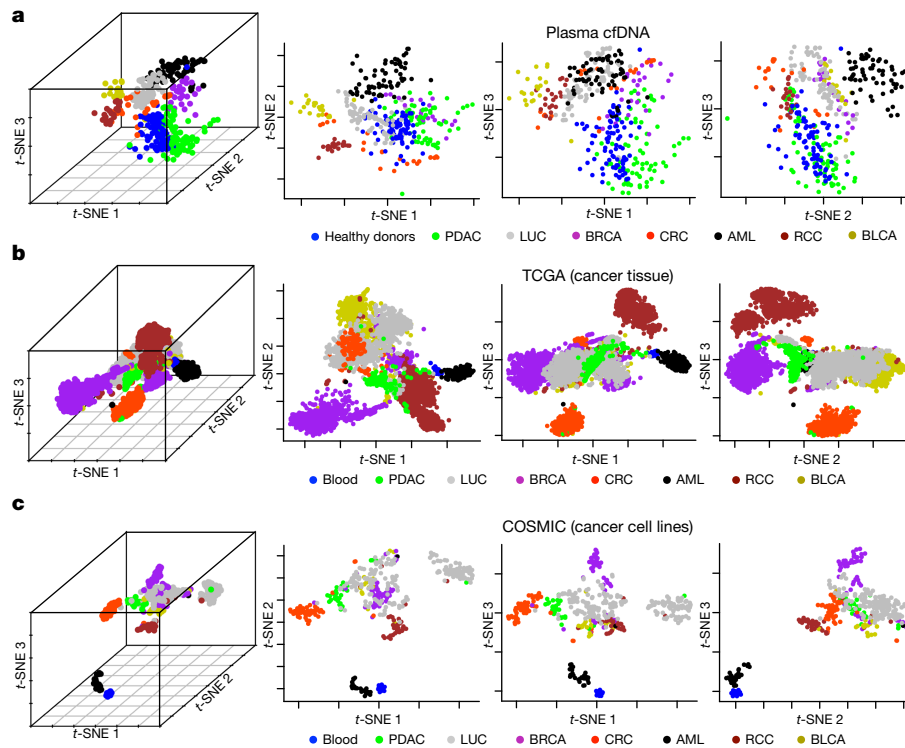


Fig. 4 | Plasma-derived DMRs are informative of cancer type. **a**, The plasma-derived DMRs identified as informative of cancer type in the discovery cohort of 189 plasma samples were used to generate 3D and 2D *t*-SNE plots for the entire cohort of plasma samples ($n = 388$). **b**, **c**, The

DNA methylation beta value for probes within the plasma-derived DMRs was used to generate 3D and 2D *t*-SNE plots for TCGA cancer tissue ($n = 4,032$) (**b**) and COSMIC cancer cell lines ($n = 400$ cell lines) (**c**).

Extended Data Fig. 6b, d, e)). Collectively, these findings indicate that cfMeDIP-seq might permit non-invasive characterization of active transcription-factor networks in cancer.

Given that we could detect tumour-specific DMRs in the plasma of PDAC cases relative to controls, we then investigated whether cfMeDIP-seq could non-invasively classify multiple cancer types from healthy controls. Consequently, we performed cfMeDIP-seq in a discovery cohort of 189 plasma samples from seven different tumour types (PDAC, CRC, breast cancer (BRCA), lung cancer (LUC), renal cancer (RCC), bladder cancer (BLCA) and acute myeloid leukaemia (AML)) and healthy controls (Extended Data Figs. 7a–l, 8a).

We first identified plasma cell-free DMRs for each tumour type relative to healthy controls. We then asked whether these cancer-type-specific DMRs identified on the plasma cfDNA were enriched for the expected tumour DMRs for each cancer type using tumour tissue methylation data from TCGA ($n = 4,032$) (Fig. 3a). We observed a marked enrichment of sites that were hypermethylated in the primary tumour tissue (TCGA) within the regions we identified as hypermethylated in the plasma cfDNA for each cancer type, coupled with significantly correlated signals between cfMeDIP-seq plasma methylation and TCGA 450k tumour data (Extended Data Fig. 8b–h). These results indicate the ability to recover ctDNA-associated methylation profiles across a range of cancer types.

Finally, we carried out a set of machine-learning analyses on our discovery cohort to rigorously evaluate the utility of cfMeDIP profiles in cancer detection and classification. We initially reduced our dataset to 505,027 windows mapping to CpG islands, shores, shelves and FANTOM5 enhancers for computational efficiency. Unbiased performance estimates, while accounting for training-set biases, were then derived from the reduced dataset. We split the discovery cohort into balanced training (80%) and test (20%) sets. Using only training-set samples, we selected the top 300 DMRs by limma-trend test statistic for each class compared with other classes. We then trained a series of one-versus-other-classes regularized binomial GLMs using these features on the training-set data. The training procedure consisted of

three rounds of 10-fold cross-validation across a grid of values for alpha and lambda with optimisation for Cohen's kappa. The use of multiple rounds of 10-fold cross-validation was motivated by a desire to leverage additional randomization for more generalizable model tuning.

The performance of these classifiers was then evaluated using receiver operating characteristic (ROC) statistics derived from the test-set samples that were not used for either DMR selection or model training. The whole process was repeated 100 times to prevent training-set biases¹⁷, culminating in a collection of 800 models, with 100 models for each one-versus-all-others comparison (hereafter termed E100). High values of the area under the receiver operator characteristic curve (AUROC) were observed for test-set samples across classes (Fig. 3b, Extended Data Fig. 9a).

Subsequently, we assessed performance across batches by applying the ensemble to a 199-sample validation cohort (35 AML, 47 PDAC, 55 LUC and 62 healthy controls). Averaging the class probabilities output by E100 for each sample yielded high AUROCs for AML versus others (0.980), PDAC versus others (0.918), LUC versus others (0.971) and normal versus others (0.969) (Fig. 3c). Notably, performance was similar between early- and late-stage samples, suggesting applicability to the detection of early-stage cancers (Fig. 3d, Extended Data Fig. 9b).

We then investigated whether the DMRs (non-zero coefficients) selected during the training of E100 were tumour-specific. Visualization using *t*-distributed stochastic neighbour embedding (*t*-SNE) plots showed clear separation by tumour type in the plasma cohort (Fig. 4a). This was notably reproduced in the 450k dataset of 4,032 TCGA cancers and normal blood samples, and 400 cancer cell lines from the Catalogue Of Somatic Mutations In Cancer (COSMIC) and PBMCs (Fig. 4b, c). This suggests that our plasma cfDNA methylation classifiers are mainly driven by tumour-specific DNA methylation patterns rather than by fluctuations in blood cells or cell composition in the tumour microenvironment.

However, these results do not rule out that some plasma cell-free DMRs could originate from changes in the proportions of circulating immune cells^{18,19}. To further test our inference, we identified 38,352

cfMeDIP windows that were lowly methylated across a range of leukocyte types in WGBS data from the International Human Epigenome Consortium (IHEC), of which 27,088 overlapped with the TCGA 450k data (Extended Data Fig. 10a). Out of these 27,088 regions, we separated those that were identified as hypermethylated through the comparisons of plasma cfDNA of each cancer type to healthy controls. We then checked the methylation status of these regions in the tumour tissue compared to PBMCs, using TCGA data for each cancer type. For PDAC, we used in-house methylation data generated for the matched patients (cfDNA and tissue DNA). We found these regions to be hypermethylated in tumour tissue (Extended Data Fig. 10b), reinforcing the hypothesis that these plasma cell-free DMRs are a direct measurement of tumour-derived DNA (that is, ctDNA).

In summary, we developed a robust, sensitive and bisulfite-free methodology for immunoprecipitation-based profiling of methylation patterns in cfDNA. Our approach awaits further validation in completely independent datasets, but our findings underscore the potential utility of cfDNA methylation profiles as a basis for non-invasive, cost-effective, sensitive and accurate early tumour detection for cancer interception, and for multi-cancer classification.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0703-0>.

Received: 5 December 2016; Accepted: 25 September 2018;

Published online 14 November 2018.

- Diaz, L. A., Jr & Bardelli, A. Liquid biopsies: genotyping circulating tumor DNA. *J. Clin. Oncol.* **32**, 579–586 (2014).
- Aravanis, A. M., Lee, M. & Klausner, R. D. Next-generation sequencing of circulating tumor DNA for early cancer detection. *Cell* **168**, 571–574 (2017).
- Newman, A. M. et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
- Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
- Phallen, J. et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **9**, eaan2415 (2017).
- Hoadley, K. A. et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).
- Lehmann-Werman, R. et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl Acad. Sci. USA* **113**, E1826–E1834 (2016).
- Visvanathan, K. et al. Monitoring of serum DNA methylation as an early independent marker of response and survival in metastatic breast cancer: TBCRC 005 prospective biomarker study. *J. Clin. Oncol.* **35**, 751–758 (2017).
- Potter, N. T. et al. Validation of a real-time PCR-based qualitative assay for the detection of methylated SEPT9 DNA in human plasma. *Clin. Chem.* **60**, 1183–1191 (2014).
- Chan, K. C. et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc. Natl Acad. Sci. USA* **110**, 18761–18768 (2013).
- Sun, K. et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc. Natl Acad. Sci. USA* **112**, E5503–E5512 (2015).
- Grunau, C., Clark, S. J. & Rosenthal, A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **29**, E65 (2001).
- Taiwo, O. et al. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protoc.* **7**, 617–636 (2012).
- Newman, A. M. et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
- Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27–36 (2010).
- Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
- Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488–492 (2005).
- Pedersen, K. S. et al. Leukocyte DNA methylation signature differentiates pancreatic cancer patients from healthy controls. *PLoS ONE* **6**, e18223 (2011).
- Teschendorff, A. E. et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS ONE* **4**, e8274 (2009).

Acknowledgements This study was conducted with support from the University of Toronto McLaughlin Centre (MC-2015-02), the Canadian Institutes of Health Research (CIHR FDN 148430 and CIHR New Investigator Salary award 201512MSH-360794-228629), Ontario Institute for Cancer Research (OICR) with funds from the province of Ontario, Canada Research Chair (950-231346), and the Princess Margaret Cancer Foundation to D.D.D.C. as well as Canadian Cancer Society (CCSRI 701717) to R.J.H., CCSRI 704716 to R.J.H. and D.D.D.C. and CCSRI 703827 to M.M.H. Recruitment of healthy individuals was supported by Cancer Care Ontario Chair of Population Health and CCSRI 020214 awarded to R.J.H. Collection of lung cancer samples was supported by the Alan B. Brown chair in molecular genomics and the Lusi Wong Lung Cancer Early Detection Program to G.L. We acknowledge the Princess Margaret Genomics Centre for carrying out the next-generation sequencing and the Bioinformatics and HPC Core, Princess Margaret Cancer Centre for their expertise in generating the next-generation sequencing data.

Reviewer information Nature thanks E. Collisson, A. Teschendorff and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.Y.S. and D.D.D.C. designed and developed the cfMeDIP-seq protocol. R.J.H. and G.F. conceived and designed the study related to the pancreatic cancer component. S.Y.S., R.S., A.C. and D.D.D.C. conceived and designed the study related to the other cancer types. S.Y.S., S.V.B., T.J.P. and D.D.D.C. designed the experiments. S.Y.S., D.C., M.H.A.R., P.C.Z., Z.C., T.L., O.K., D.R., I.E., Z.C., S.C., G.M.O., J.L., M.M. and Z.Z. performed the experiments. T.d.S.M., Y.W. and C.O. performed the mouse experiments. R.S., A.C., G.F., T.T.W., A.G., T.J.P., M.M.H. and D.D.D.C. analysed the data with scientific input from R.J.H. G.F., A.B., D.C., A.S., T.M., A.A., N.L., M.H.A.R., J.D.M., P.L.B., N.F., G.L., M.D.M., S.G., T.J.P. and R.J.H. collected the clinical data related to the samples, determined the sample selection criteria and matching scheme, and provided the clinical samples. S.Y.S., R.S., A.C. and D.D.D.C. wrote the paper with feedback from all authors.

Competing interests D.D.D.C., S.Y.S., A.C., S.V.B., R.S. and R.J.H. are listed as inventors/contributors on patents filed related to this work.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0703-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0703-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to R.J.H. or D.D.D.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. Plasma samples were blinded during the sample preparation and sequencing. Data analysis was performed unblinded on the discovery cohort and blinded on the validation cohort.

Bioinformatic simulation of tumour-specific features and probability of detection by sequencing depth. We created 145,000 simulated genomes with 1, 10, 100, 1,000 and 10,000 independent loci with 0.001–10% cancer-specific DMRs in tenfold increments. Diploid genomes (14,500, the expected copy number in 100 ng cfDNA) were then sampled from these mixtures and further sampled 10–10,000× in tenfold increments at each locus. The process was repeated 100 times for each combination of parameters. Probability curves were plotted for successful detection of >1 and >5 DMRs (Fig. 1a, Extended Data Fig. 1a).

cfMeDIP-seq. A schematic representation of the cfMeDIP-seq protocol is shown in Extended Data Fig. 1b. Before cfMeDIP, the samples were subjected to library preparation using Kapa HyperPrep Kit (Kapa Biosystems), following the manufacturer's protocol with minor modifications. In brief, after end-repair and A-tailing, samples were ligated to 0.181 μM of NEBNext adaptor (NEBNext Multiplex Oligos for Illumina kit, New England BioLabs) by incubating at 20 °C for 20 min and purified with AMPure XP beads (Beckman Coulter). The eluted library was digested using the USER enzyme (New England BioLabs) followed by purification with Qiagen MinElute PCR Purification Kit (MinElute columns) before MeDIP.

The prepared libraries were combined with the filler λ DNA (to ensure the total amount of DNA (cfDNA + filler) was 100 ng) and subjected to MeDIP with Diagenode MagMeDIP kit (C02010021) using a previously published protocol¹³ with some modifications. The filler DNA consists of a mixture of unmethylated and in vitro methylated λ amplicons of different CpG densities (Supplementary Table 6), similar in size to adaptor-ligated cfDNA libraries. Its addition ensures a constant ratio of antibody to input DNA and helps to maintain similar immunoprecipitation efficiency across samples regardless of available cfDNA, while minimizing non-specific binding by the antibody and DNA loss due to binding to plasticware. For MeDIP, the prepared library/filler DNA mixture was combined with 0.3 ng of control methylated and 0.3 ng of the control unmethylated *Arabidopsis thaliana* DNA provided in the kit, and the buffers. The mixture was heated to 95 °C for 10 min, then immediately placed into an ice water bath for 10 min. Each sample was partitioned into two 0.2 ml PCR tubes: one for the 10% input control (7.9 μl) and the other for the sample to be subjected to immunoprecipitation (79 μl). The included 5-mC monoclonal antibody 33D3 (C15200081) from the MagMeDIP kit was diluted 1:15 before generating the diluted antibody mix and was added to the sample. Washed magnetic beads (following the manufacturer's instructions) were also added before incubation at 4 °C for 17 h. The samples were purified using the Diagenode iPure Kit v2 (C03010015) and eluted in 50 μl of buffer C. The success of the reaction (QC1) was validated by qPCR to detect recovery of the spiked-in methylated and unmethylated *A. thaliana* DNA. The percentage recovery of unmethylated spiked-in DNA should be <1% (relative to input control, adjusted for input control being 10% of the overall sample) and the percentage specificity of the reaction should be >99% (as calculated by $(1 - [\text{recovery of spiked-in unmethylated control DNA over recovery of spiked-in methylated control DNA}]) \times 100$), before proceeding to the next step. The optimal number of cycles to amplify each library was determined by qPCR, after which the samples were amplified using Kapa HiFi Hotstart Mastermix and NEBNext multiplex oligos, added to a final concentration of 0.3 μM. The final libraries were amplified as follows: activation at 95 °C for 3 min, followed by predetermined cycles of 98 °C for 20 s, 65 °C for 15 s and 72 °C for 30 s and a final extension of 72 °C for 1 min. The amplified libraries were purified using MinElute columns, then gel size selected with 3% Nusieve GTG agarose gel to remove any adaptor dimers. All the final libraries were submitted for BioAnalyzer analysis before sequencing at the Princess Margaret Genomics Centre on an Illumina HiSeq 2500, SBS V4 chemistry, single read 50 bp, multiplexed as seven samples per lane. After sequencing, the sequenced reads were aligned to λ and hg19 using Bowtie²⁰ with the default settings. On the basis of virtually no alignment to the λ genome, the filler DNA does not interfere with the generation of sequencing data (Supplementary Tables 7, 8).

The generated SAM files from hg19 alignment were converted to BAM format, ensuring the removal of duplicate reads, and the reads were then sorted and indexed using SAMtools²¹ before subsequent analysis with the R package MEDIPS²². The CpG enrichment score, as a quality control measure for the immunoprecipitation reaction, was calculated as part of the MEDIPS package.

Validation of cfMeDIP-seq against MeDIP-seq. DNA from human colorectal cancer cell (CRC) line HCT116 (American Type Culture Collection (ATCC), STR tested for authentication, mycoplasma free) was extracted using PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific). HCT116 was chosen because of the availability of public DNA methylation data. Genomic DNA was sheared to mimic cfDNA using a Covaris sonicator, and larger size fragments were excluded using AMPure XP beads (Beckman Coulter) to mimic the fragment size of cell-free

DNA. cfMeDIP-seq was carried out on 1, 5, 10 and 100 ng of sheared DNA as input, with 100 ng representing the gold-standard MeDIP-seq protocol, with two biological replicates per input. The fold enrichment of a methylated human DNA region (*HIST1H2BA*) over unmethylated human DNA region (*GAPDH* promoter), using primers provided in the MagMeDIP kit, was determined before sequencing libraries to saturation (Extended Data Fig. 2a–c, Supplementary Table 7).

Dilution series of sheared cell line DNA. As with the CRC DNA, the same extraction and shearing protocol was used with multiple myeloma cell line MM.1S (source: American Type Culture Collection (ATCC), STR tested for authentication, mycoplasma free). A dilution series of CRC into multiple myeloma DNA was carried out following the scheme in Extended Data Fig. 3a. This dilution series was used for cfMeDIP-seq (Supplementary Table 9) and for ultra-deep targeted sequencing for CRC point-mutation detection, using a starting input of 60 ng of DNA. For the mutation detection, DNA libraries were prepared using Kapa HyperPrep Kit (Kapa Biosystems) and Illumina compatible molecular barcoded adapters with 2-bp in-line barcodes (unique molecular identifiers (UMIs)) to ensure optimal analytical sensitivity for mutation detection¹⁴. A customized biotinylated DNA capture probe panel (xGen Lockdown Custom Probes Mini Pool, Integrated DNA Technologies) targeting exons from five genes (13 kb) was used²³. In brief, the barcoded libraries were pooled, and hybrid capture was performed according to the manufacturer's instructions (IDT xGen Lockdown protocol version 4). The amplified post-capture libraries were sequenced to >100,000× read coverage using Illumina HiSeq 2500 instrument, SBS V4 chemistry, paired-end 125 bp, as four samples per lane. Average target coverage of unprocessed reads was $186,312 \times$ (range: $154,419 \times - 216,434 \times$) (Supplementary Table 9).

After sequencing, reads were de-multiplexed using sample-specific indices into separate paired-end FASTQ files. A two-base-pair molecular barcode and a one-base-pair invariant spacer sequence were removed from each read. A thymine base was encoded in the third position for adaptor ligation and a spacer filter was enforced to remove reads that were incompatible with this design. The extracted barcodes from paired-end reads were grouped and written into the header of each sequence for downstream in silico molecular identification²⁴. FASTQ files were mapped to the human reference genome hg19 using BWA²⁵, processed using the Genome Analysis ToolKit (GATK) IndelRealigner²⁶, and sorted and indexed using SAMtools²¹.

Barcodes were used in combination with endogenous sequence features (genome coordinates, mapping alignments, read orientation, and read number in pair) to confer sequences from individual molecules. Consensus sequences were formed from two or more reads supporting the same molecule with 70% agreement amongst bases above Phred quality scores²⁷ (Q) of 30. Reads derived from the same strand of a unique fragment were collapsed to form SSCs, suppressing polymerase and sequencer errors. These condensed reads were subsequently combined with their complementary strand into DCSs. This enables an additional layer of error suppression as double-strand consolidated sequences can correct for asymmetric damage accrued during the first cycle of PCR or induced by oxidation²⁸.

We selected variants on the basis of annotated SNPs from the Cancer Cell Line Encyclopedia²⁹ overlapping our target panel. SNVs were called with MuTect³⁰ using the following parameters: `-enable_extended_output-tumor_f-pretest 0.000001f-downsampling_type NONE-force_output-force_alleles-gap_events_threshold 1000-fraction_contamination 0.00f-coverage_file30`. We force called every base for each variant to assess limit of detection and background noise at each stage of barcode-mediated error correction. Analysis of the UMI-processed error-suppressed reads revealed unique molecule (that is, SSCs) and DCS average target coverage of $6,276 \times$ ($4,284 \times - 8,068 \times$) and $1,043 \times$ ($654 \times - 1,602 \times$), respectively (Supplementary Table 9).

Specimen processing of patient-derived xenograft cfDNA. All mouse work was carried out in compliance with animal use protocol and ethical regulations approved by the Animal Care Committee at University Health Network (UHN). Human colorectal tumour tissue obtained with patient consent and UHN Research Ethics Board approval from the UHN Biobank was digested to single cells using collagenase A. Single cells were subcutaneously injected into 4–6-week-old NOD/SCID male mice. Mice were euthanized by CO₂ inhalation before blood was collected by cardiac puncture and stored in EDTA tubes. From the collected blood samples, plasma was isolated and stored at –80 °C. cfDNA was extracted from 0.3–0.7 ml of plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen). Two biological samples with 10 ng of starting cfDNA were subjected to the cfMeDIP-seq protocol as previously mentioned, sequenced and analysed (Supplementary Table 10).

Donor recruitment and sample acquisition. All patients provided written informed consent, and all samples were obtained upon approval of the institutional ethics committees and Research Ethics Boards from UHN and Mount Sinai Hospital, in compliance with all relevant ethical regulations. Pancreatic adenocarcinoma cases were obtained from the Ontario Pancreatic Cancer Study and the UHN Biobank. Colorectal and breast cancer plasma samples were obtained

from the UHN Biobank. Lung cancer plasma samples were obtained from the UHN Thoracic Biobank. AML samples were obtained from the UHN Leukaemia Biobank. Bladder and renal cancer plasma samples were obtained from the UHN Genitourinary Biobank from consenting urologic oncology patients, procured before nephrectomy and cystectomy respectively. Healthy controls were recruited through the Family Medicine Centre at Mount Sinai Hospital in Toronto, Canada. **Specimen processing and methylation analysis of purified tumour and normal cells from PDAC samples.** For primary PDAC samples, specimens were processed immediately following resection and representative sections were used to confirm the diagnosis. Laser capture microdissection (LCM) of freshly liquid nitrogen-frozen tissue samples was performed on a Leica LMD 7000 instrument. Laser capture microdissection was performed on the same day as sections were cut to minimize nucleic acid degradation. Qiagen Cell Lysis Buffer was used to extract genomic DNA.

Quantified 10 ng of genomic DNA for each sample was analysed using RRBS following a previously published protocol³¹ with minor modifications. DNA libraries ligated to Illumina TruSeq methylated adapters were subjected to bisulfite conversion using the Zymo EZ DNA methylation kit following the manufacturer's protocol, followed by gel size selection for fragments of 160–300 bp in size. After determining the optimal number of cycles to amplify each purified library, samples were amplified using Kapa HiFi Uracil+ Mastermix (Kapa Biosystems) and purified with AMPure beads (Beckman Coulter). The final libraries were submitted for BioAnalyzer analysis before sequencing at the Princess Margaret Genomics Centre on an Illumina HiSeq 2000, using sequencing by synthesis (SBS) V3 chemistry, single read 50 bp and multiplexed as four samples per lane. After sequencing, the raw data for each sample was trimmed with Trim Galore! using the RRBS settings before aligning to hg19 using Bismark³² with Bowtie2³³ (Supplementary Table 11). The generated SAM files were then converted to BAM format, sorted and indexed using SAMtools.

Specimen processing for patient cfDNA. Plasma samples collected using EDTA and acid citrate dextrose tubes were obtained from the UHN BioBanks and Mount Sinai Hospital and were kept frozen until use. cfDNA was extracted from 0.5–3.5 ml of plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) and quantified through Qubit before use. The sex, age and pathology stage of the patients from which the samples were collected are available in Supplementary Table 12, and extracted DNA quantities are available in Extended Data Fig. 8a.

Calculation and visualization of differentially methylated regions from cfDNA of patients with pancreatic cancer and healthy donors. DMRs between cfDNA samples from 24 patients with pancreatic cancer (PDAC) and 24 healthy donors (controls) were calculated using MEDIPS and DESeq2 R packages^{22,34}. For each sample, we computed counts per 300 bp non-overlapping windows, filtered out windows with less than 10 counts across all samples and fit a negative binomial model to call DMRs at FDR < 0.1 (Wald test). *z*-scores of DMR RPKM values with Euclidean distance and Ward clustering were used for visualization.

Enrichment analyses for plasma-derived DMRs in tumour-specific methylation signals in PDAC. Five normal PBMC samples profiled by RRBS were downloaded from the Gene Expression Omnibus (accession number GSE89473) for comparison with the 24 pancreatic cancer tissue RRBS samples. The R package MethylKit was used to parse files and autosomal CpGs detected in at least 18 out of the 24 PDACs and 4 out of the 5 PBMCs were retained for further analysis. We obtained DMCs at FDR < 0.01, delta beta > 0.25. A null distribution was then generated from 1,000 resamples, preserving the relationship between the number of CpGs in windows that were seen in the original intersections between RRBS features and cfMeDIP DMRs. Then we computed the frequency of overlap between DMRs hypermethylated in both, hypermethylated in one but not the other, hypomethylated in one but not the other, and finally, hypomethylated in both comparisons. The distributions were then standardized based on *z*-scores and used to compute Bonferroni-adjusted *P* values to determine enrichment. The same procedure was employed for subsequent enrichment tests in the manuscript.

Enrichment analyses for cfMeDIP DMRs in TCGA 450K DMCs relative to normal tissues and PBMCs. 189 cfDNA samples were obtained across seven cancer types (AML, bladder (BLCA), breast (BRCA), colorectal (CRC), lung (LUC), pancreatic (PDAC) and renal cancer (RCC)) and healthy donors (normal) (Supplementary Table 12). After processing of cfMeDIP-seq data from these samples, DMRs were calculated using DESeq2 between each cancer type and healthy donors as described above. DMCs were also calculated between TCGA 450K methylation array samples from each corresponding cancer type (*n* = 3,979) (obtained from SAGE synapse) and PBMCs (*n* = 53, obtained from the Gene Expression Omnibus) samples using limma (FDR < 0.01, absolute delta beta 0.25). Statistical tests for enrichment were performed as described above for PDAC RRBS samples. The same procedure was carried out for DMCs calculated between TCGA 450K methylation array samples from a cancer type and normal samples from the same tissue, for BLCA, BRCA, CRC, LUC and RCC.

Examination of transcription factors associated with differentially methylated motifs in cfMeDIP-seq DMRs. RNA-seq data obtained as median RPKMs from the GTEx consortium across 53 human tissues—as described in the supplementary R Markdowns in Zenodo (ID 10.5281/zenodo.1205756) (Supplementary Table 13)—and median expression per tissue was visualized in heat maps. To look for enrichment of transcription factor expression and DMR-associated transcription factor motifs, we selected 1,000 random sets of transcription factors. As part of the analysis, we considered the known sensitivity to the methylation status of each transcription factor¹⁶, yielding 42 transcription factors that are enriched in healthy donors and 52 that are enriched in pancreatic adenocarcinoma cases.

We computed ssGSEA (single-sample gene set enrichment analysis) scores for the expression of these transcription factors per sample, for pancreatic cancer (TCGA), blood (GTEx) and normal pancreas (GTEx) and compared distributions to those from random sets of transcription factors using Wilcoxon's Rank Sum Test. Violin plots were constructed as described in the supplementary R Markdown 10.5281/zenodo.1205735 (Supplementary Table 13).

Machine learning analyses for evaluation of classification accuracy. *Model training and evaluation on the discovery cohort.* In order to evaluate the performance of cfMeDIP data in tumour classification without high computational cost, we reduced the initial set of possible candidate features to windows encompassing CpG islands, shores, shelves and FANTOM5 enhancers ('regulatory features'), yielding a matrix of 189 samples and 505,027 features.

We then used the caret R package³⁵ to partition the discovery cohort data into 100 class-balanced independent training and test sets in an 80–20% manner. Then, we selected the top 300 DMRs by moderated *t*-statistic (150 hypermethylated, 150 hypomethylated) on the training data partition using limma-trend³⁶ for each class versus other classes. A binomial GLMnet was then trained using these DMRs (up to 300 DMRs × 7 other classes = 2,100 features) using three iterations of 10-fold cross-validation to optimize values of the mixing parameter (alpha, values = 0, 0.2, 0.5, 0.8 and 1) and the penalty (lambda, values = 0–0.05 in increments of 0.01) using Cohen's Kappa as the performance metric. For each training set, this yielded a collection of eight one-class-versus-other-classes binomial classifiers.

We then estimated classification performance on the held-out test set using the AUROC (area under the receiver operating characteristic curve). These estimates represent unbiased measures of classification, as the held-out test set samples were not used for either DMR pre-selection or GLMnet training and tuning. The 100 independent training and test sets also permitted the minimization of optimistic estimates owing to training-set bias.

Model evaluation on the validation cohort. For each validation cohort cfMeDIP sample, we estimated class probabilities for the AML, PDAC, LUC and normal one-versus-all binomial classifiers trained on the 100 different training sets within the discovery cohort. The probabilities from the 100 models were averaged to produce a single score that was then used for AUROC estimation. We also evaluated if disease stage (applicable to only LUC and PDAC) affected performance by estimating AUROC when either early- (stages I and II) or late-stage samples (stages III and IV) of a particular class were left out for the one-versus-all classifiers trained to identify the class in question.

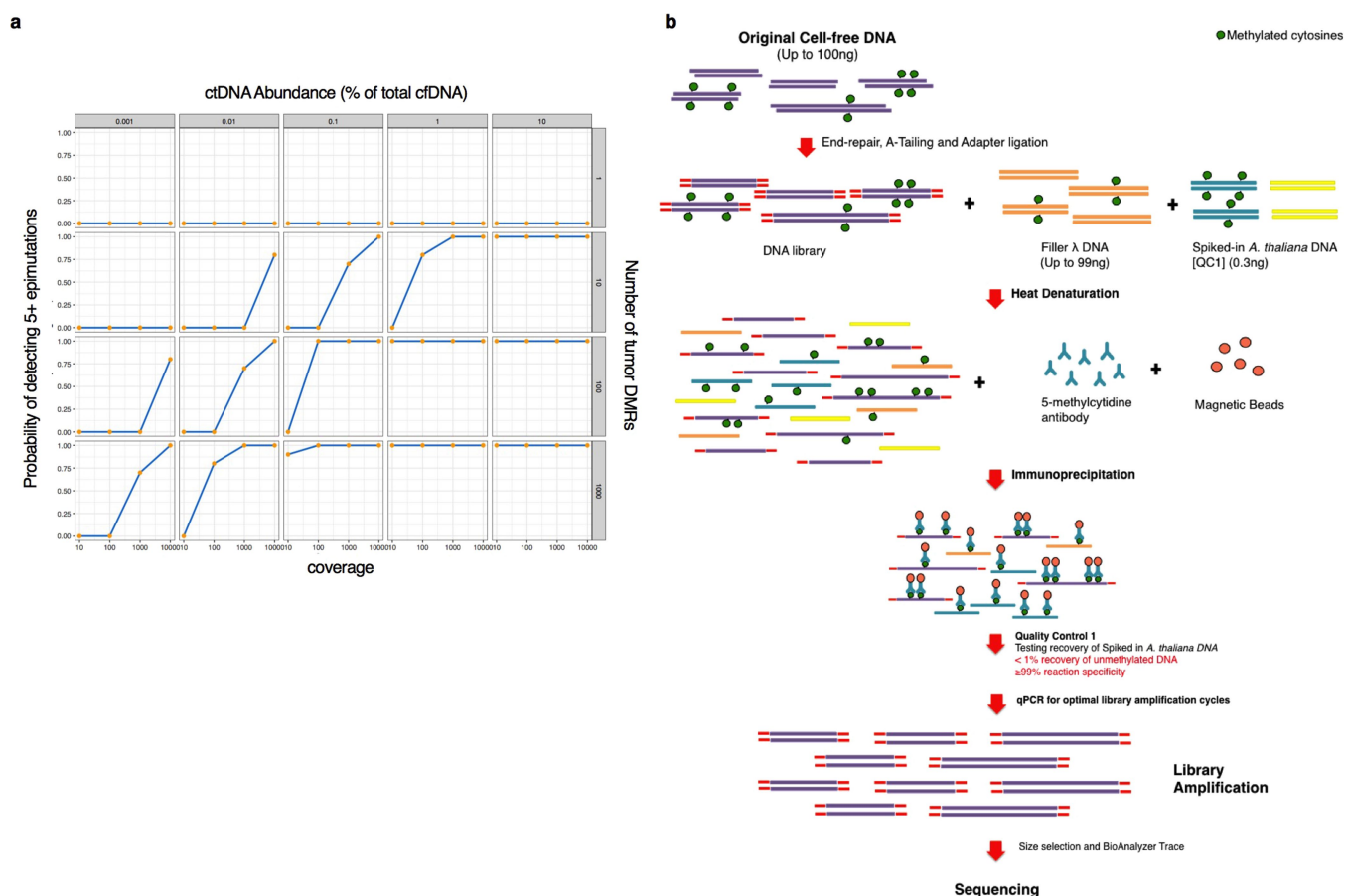
Validation in cell lines. 450K profiles for 1,028 cell lines previously characterized³⁷ were obtained as IDAT files. The data were then uniformly processed using the ssNoob method in the minfi package³⁸. We reduced this dataset to tissue types for which cfMeDIP data were available (*n* = 400).

Data availability

R markdowns (either knit or raw) and scripts used to generate the findings in this study have been deposited on Zenodo (DOIs in Supplementary Table 13). All the cell line datasets generated and/or analysed during the current study are available in the Gene Expression Omnibus repository under accession code GSE79838. The cfMeDIP-seq next-generation sequencing data for patient samples that support the findings of this study are available upon request from the corresponding author to comply with institutional ethics regulation. Source data for Fig. 1b and Extended Data Fig. 3e are provided in Supplementary Table 9, and for Fig. 1c in Supplementary Table 10. Additional source data can be found on Zenodo (Supplementary Table 13).

- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Lienhard, M., Grimm, C., Morkel, M., Herwig, R. & Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**, 284–286 (2014).
- Kis, O. et al. Circulating tumour DNA sequence analysis as an alternative to multiple myeloma bone marrow aspirates. *Nat. Commun.* **8**, 15086 (2017).

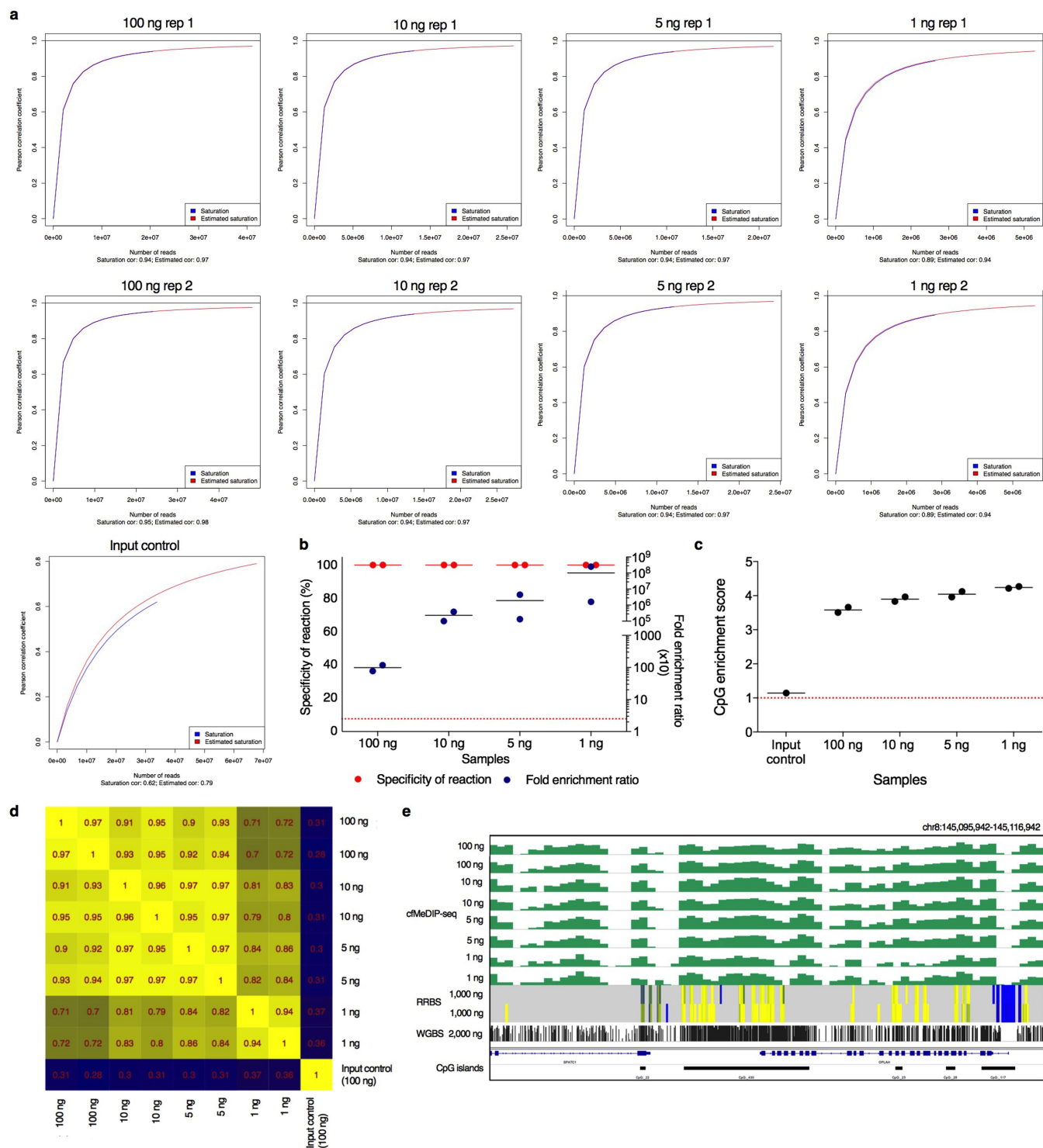
24. Kennedy, S. R. et al. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protoc.* **9**, 2586–2606 (2014).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
26. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
27. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
28. Schmitt, M. W. et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109**, 14508–14513 (2012).
29. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
30. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
31. Gu, H. et al. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* **6**, 468–481 (2011).
32. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
33. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
35. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, (2008).
36. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
37. Iorio, F. et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).
38. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).



Extended Data Fig. 1 | Simulation of the probability of detecting ctDNA as a function of the number of DMRs, sequencing depth and percentage of ctDNA in plasma cfDNA, and a proposed method to enrich ctDNA.

a, Bioinformatic simulation of scenarios with different proportions of ctDNA present in the sample (0.001% to 10%, columns), and a range of tumour-specific DMRs—from 1, 10, 100, 1,000 or 10,000—determined through the comparison of ctDNA to normal cfDNA (rows), with

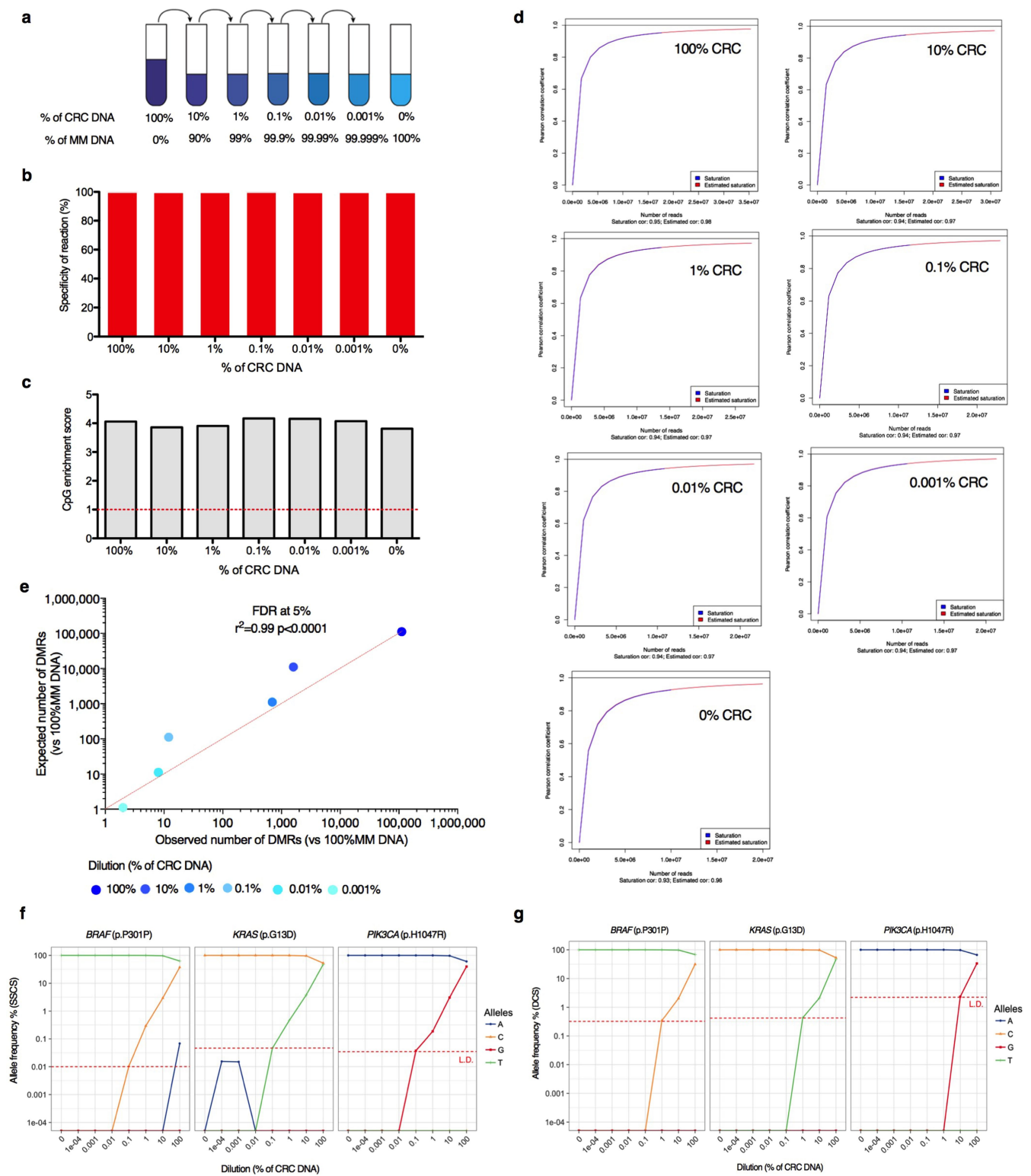
reads sampled at varying sequencing depths at each locus (10×, 100×, 1,000× and 10,000×) (*x* axis). The probability of detecting at least five epimutations per DMR increases as the number of available features increases, even at shallow coverage per locus (left *y* axis). Each panel depicts probability of detection against coverage per candidate DMR for one simulation scenario. **b**, Schematic representation of the cfMeDIP-seq protocol.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Sequencing saturation analysis and quality controls of MeDIP-seq and cfMeDIP-seq carried out on varying starting inputs of HCT116 DNA sheared to mimic cfDNA. **a**, Results of the saturation analysis from the Bioconductor package MEDIPS analysing cfMeDIP-seq data from each replicate, for each starting input amount and including an input control. **b**, The protocol was tested in two biological replicates of four starting DNA inputs (100, 10, 5 and 1 ng) of HCT116 DNA sheared to mimic cfDNA. The specificity of the reaction was calculated using methylated and unmethylated spiked-in *A. thaliana* DNA. The fold-enrichment ratio was calculated using genomic regions of the fragmented HCT116 DNA (human methylated *HIST1H2BA* and unmethylated *GAPDH*). The horizontal dotted line indicates a fold-enrichment ratio threshold of 25, dots represent biological replicates, with lines representing the mean. **c**, CpG enrichment scores of the sequenced samples (two biological replicates each of four starting DNA inputs (100, 10, 5 and 1 ng) and one input control) show a robust enrichment of CpGs within the genomic regions from the immunoprecipitated samples

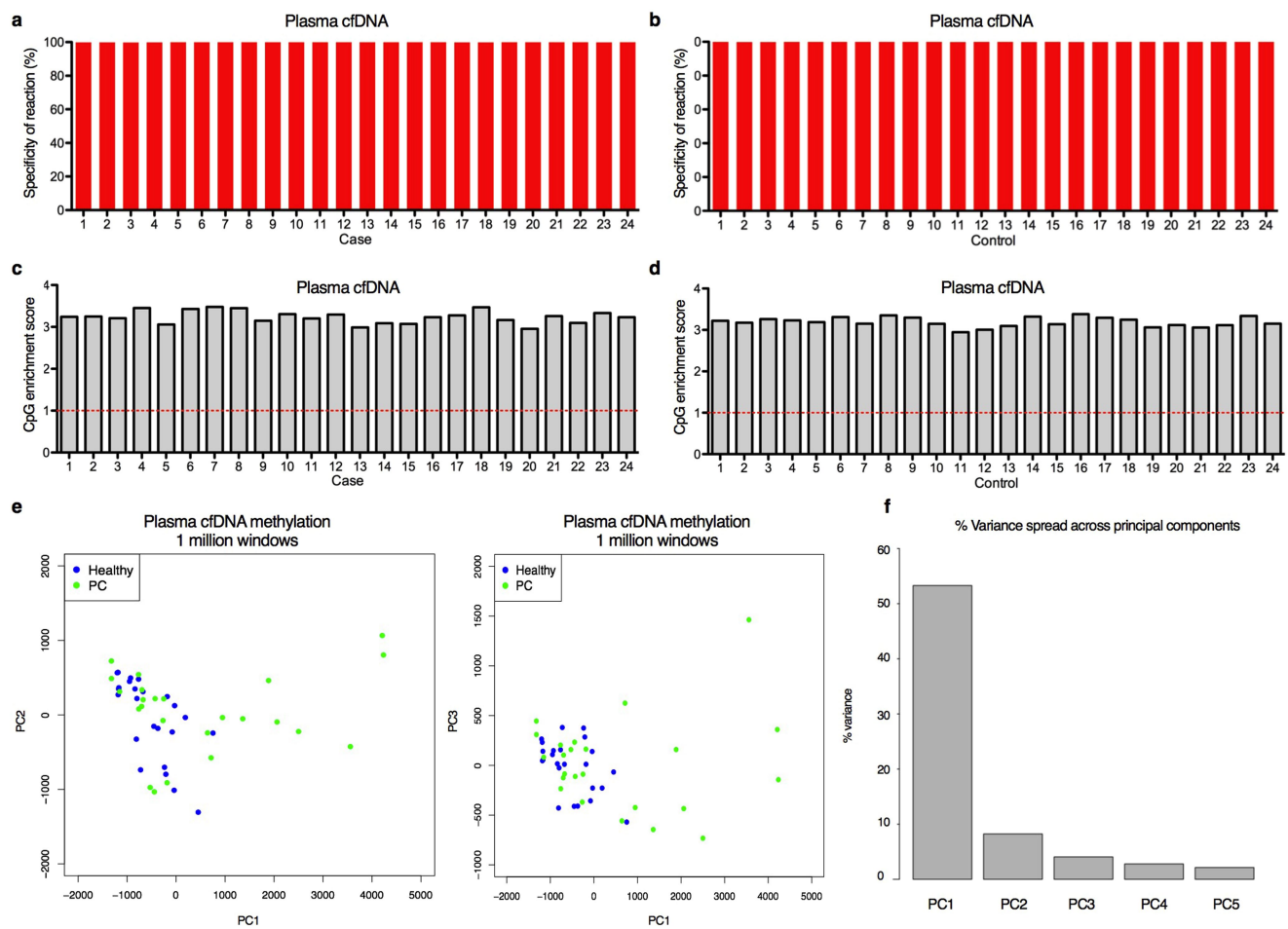
compared to the input control. The CpG enrichment score was obtained by dividing the relative frequency of CpGs of the regions by the relative frequency of CpGs of the human genome. The horizontal dotted line indicates a CpG enrichment score of 1, dots represent biological replicates, with lines representing the mean. **d**, Genome-wide Pearson correlations of normalized read counts per 300-bp window between cfMeDIP-seq signal for 1 to 100 ng of input HCT116 DNA sheared to mimic cfDNA (2 biological replicates per concentration). **e**, Genome Browser snapshot of HCT116 cfMeDIP-seq signal across a window (chr8:145,095,942–145,116,942) selected out of four examined loci, at different starting DNA inputs (1 to 100 ng, in biological replicates), compared with RRBS (ENCODE: ENCSR000DFS) and WGBS (Gene Expression Omnibus: GSM1465024) data (aligned to hg19). For cfMeDIP-seq, the y axis indicates RPKMs; for RRBS, yellow and blue blocks represent hypermethylated and hypomethylated CpGs, respectively. In the WGBS track, peak heights indicate methylation level.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Sequencing saturation analysis and quality controls of cfMeDIP-seq from serial dilution. **a**, Schematic representation of the CRC DNA (HCT116) dilution series into multiple myeloma DNA (MM.1S). For both CRC and multiple myeloma DNA, the genomic DNA was sheared to mimic cfDNA fragmentation. The entire dilution series was used to carry out cfMeDIP-seq ($n = 1$) and ultra-deep sequencing for mutation detection ($n = 1$). **b**, The specificity of the reaction for each dilution in the series ($n = 1$) was calculated using methylated and unmethylated spiked-in *A. thaliana* DNA. **c**, CpG enrichment representing the ratio of relative frequency of CpGs in regions to relative frequency of CpGs in the human genome for each dilution in the series ($n = 1$), determined by cfMeDIP-seq. The horizontal dashed line represents a CpG enrichment of 1. **d**, Saturation analysis of cfMeDIP-seq

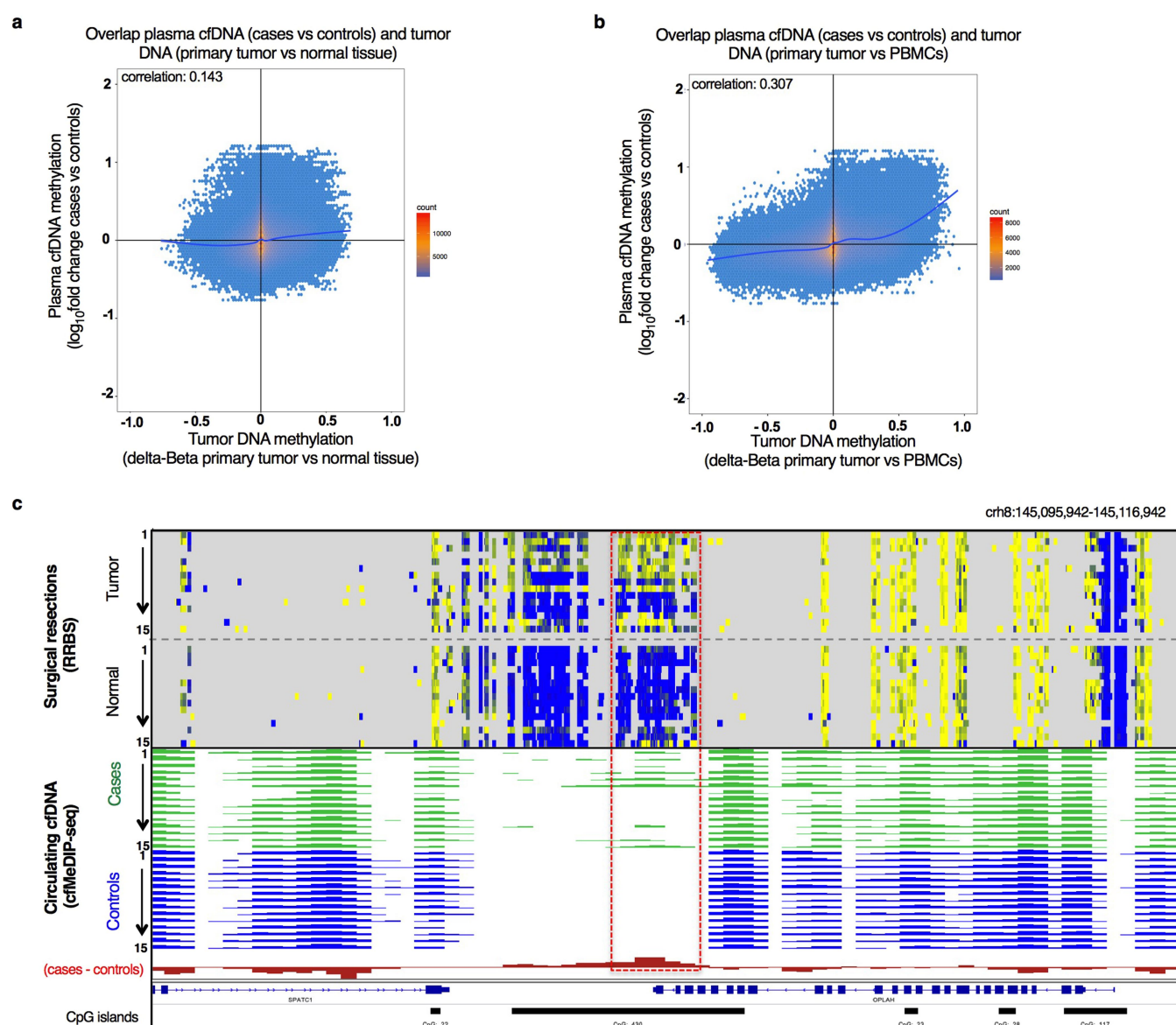
sequenced reads from each dilution point in the series ($n = 1$). **e**, Across a serial dilution series ($n = 7$ dilution points, two technical replicates, each replicate was used per protocol) of HCT116 DNA spiked into MM.1S multiple myeloma DNA, near-perfect correlations are observed between observed and expected numbers of DMRs. **f**, **g**, Ultra-deep sequencing for mutation detection of three CRC-specific point mutations within *BRAF* (p.P301P), *KRAS* (p.G13D) and *PIK3CA* (p.H1047R) in the same dilution series (of CRC into multiple myeloma DNA) ($n = 1$). UMIs were incorporated into the sequencing adapters and used to create SSCs (**f**) and DCSs (**g**) for the detection of allele frequency for each mutation at each locus. For each mutation, the reference allele is found at the top. The dashed red line indicates the limit of detection.



Extended Data Fig. 4 | Quality control of cfMeDIP-seq from circulating cfDNA from patients with PDAC (cases) and healthy donors (controls).

a, b, Specificity of reaction calculated using methylated and unmethylated spiked-in *A. thaliana* DNA for each case sample (**a**) and each control sample (**b**). The fold-enrichment ratio was not calculated owing to the very limited amount of DNA available after final libraries were generated. **c, d**, CpG enrichment of the sequenced cases (**c**) and controls (**d**). The

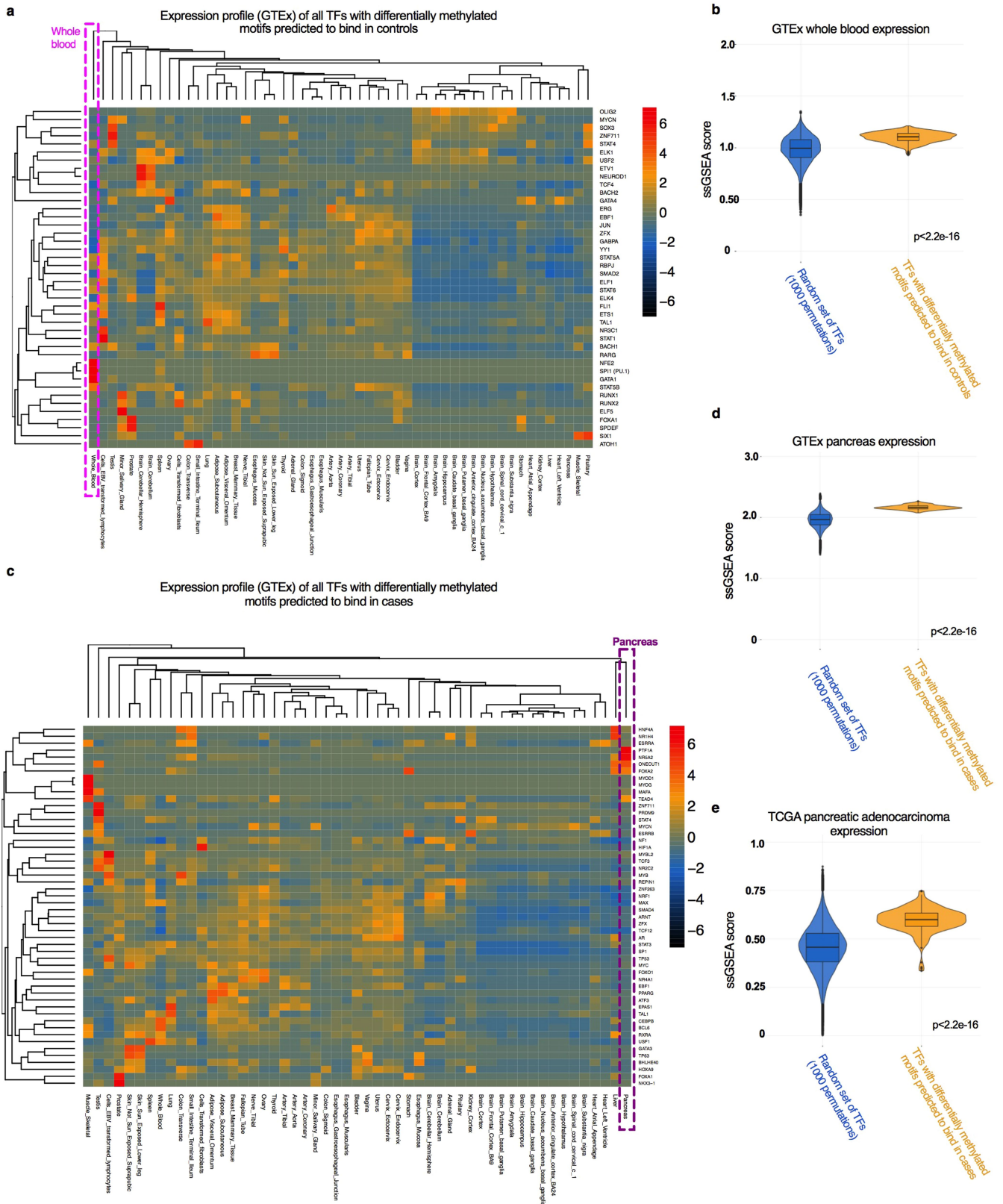
horizontal dashed line represents a CpG enrichment of 1. **e**, Principal component (PC) analysis of cfDNA methylation from 24 plasma cfDNA samples from healthy donors and 24 plasma cfDNA samples from patients with PDAC, using the 1 million most variable windows by median absolute deviation (300 bp) genome-wide. Left, PC2 against PC1; right, PC3 against PC1. **f**, Percentage of variance explained by each principal component.



Extended Data Fig. 5 | Methylome analysis of plasma cfDNA distinguishes patients with early-stage PDAC from healthy controls.

a, The difference in plasma cfDNA methylation plotted against the difference in tumour DNA methylation for each overlapping window ($n = 547,887$). The difference in plasma cfDNA methylation between patients with PDAC and healthy controls is \log_{10} -fold, as measured by cfMeDIP-seq. Tumour DNA methylation difference is delta beta from primary PDAC tumour to normal tissue, as measured by RRBS. The blue line is a trend line, with the correlation determined by Pearson's correlation. **b**, Scatter plot showing the DNA methylation difference for each overlapping window. The x axis shows the DNA methylation difference for the primary PDAC tumour compared with normal PBMCs from the RRBS data. The y axis shows the DNA methylation difference for the plasma cfDNA methylation from patients with PDAC compared

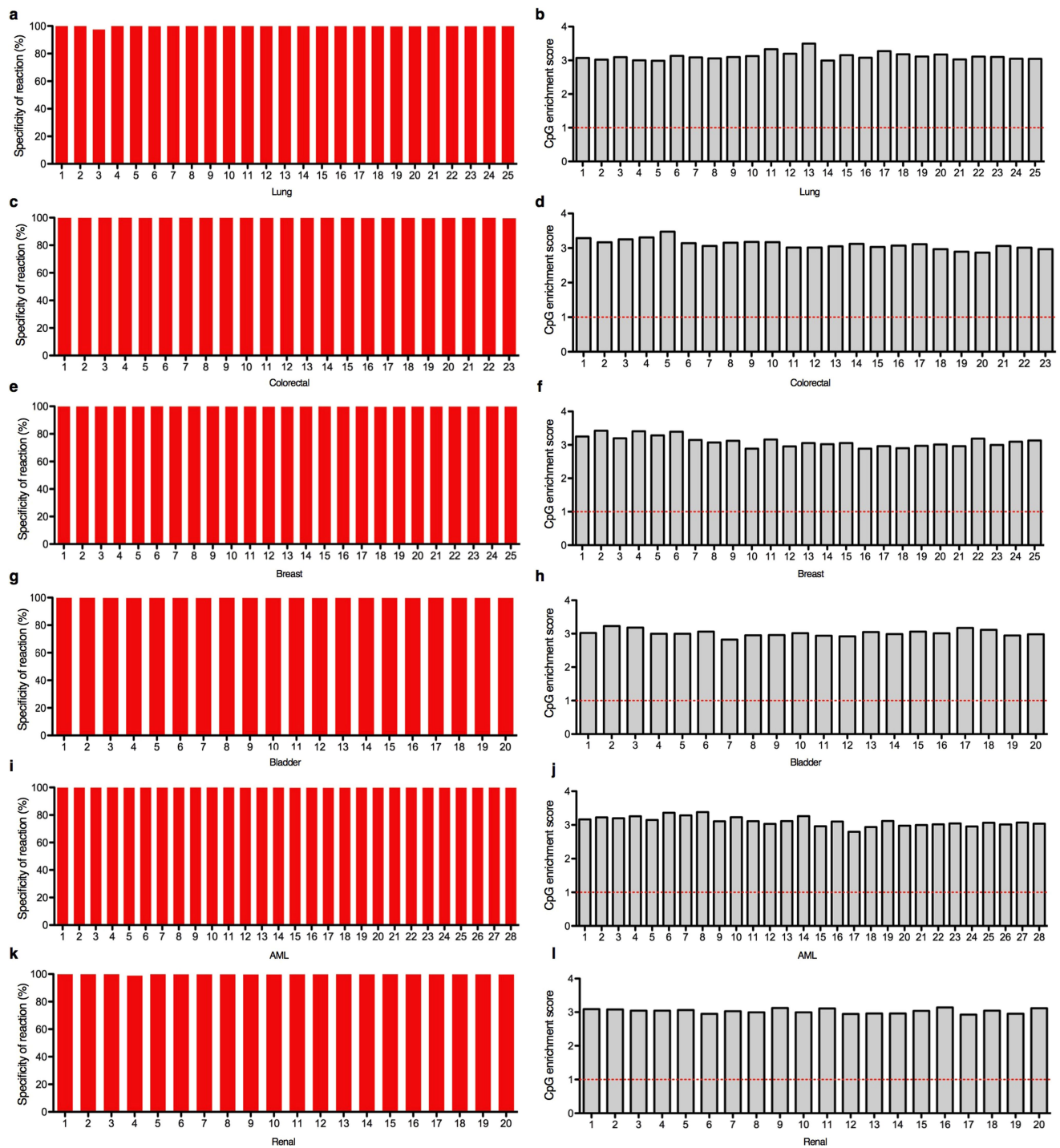
with healthy donors from the cfMeDIP-seq data. Correlation determined by Pearson's correlation. **c**, Genome Browser snapshot of RRBS and cfMeDIP-seq signal across a representative chromosomal region selected from four candidate regions (chr8:145,095,942–145,116,942) using reference genome hg19. RRBS tracks show the methylation signal for the laser capture microdissection tissues from PDAC tumour cases and the matching normal tissue, from the same patient, shown in the same order. Each coloured block represents DMCs, with yellow representing hypermethylated and blue representing hypomethylated. cfMeDIP-seq tracks show the methylation signal (RPKMs) detected in the cfDNA, with cases representing plasma from the same PDAC cases and controls corresponding to plasma from age- and sex-matched healthy controls. For the cfMeDIP-seq tracks, green and blue peaks indicate the methylation signal (RPKMs) detected in the cfDNA.



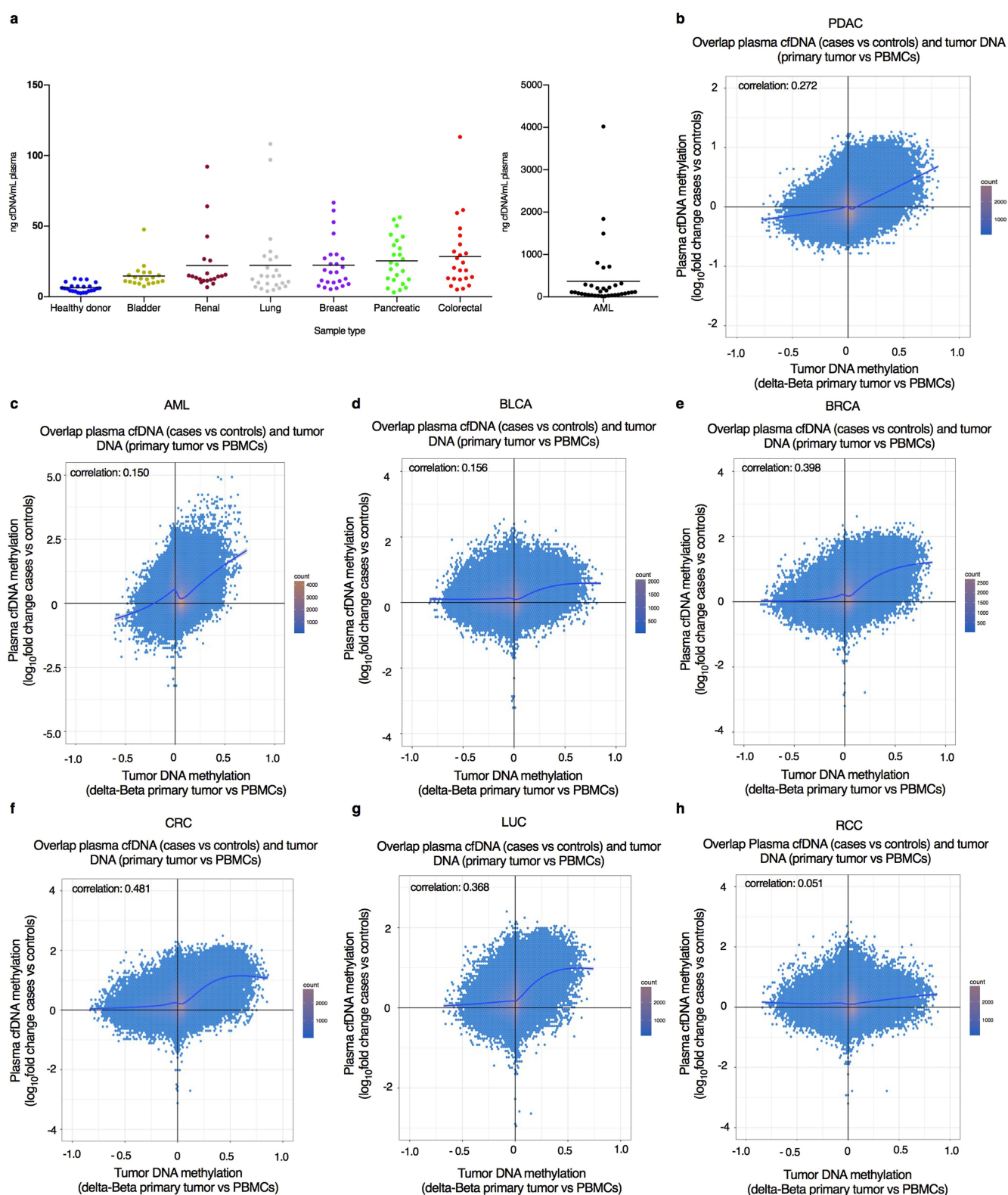
Extended Data Fig. 6 | See next page for caption

Extended Data Fig. 6 | Circulating cfDNA methylation profiles can identify transcription factor footprints and infer active transcriptional networks in the tissue of origin. **a**, Expression profile of all transcription factors ($n = 42$) that were characterized as binding in healthy controls across 53 human tissues from the GTEx project. Several transcription factors that are preferentially expressed in the haematopoietic system were identified (PU.1, NFE2 and GATA1). **b**, Expression profiles (ssGSEA scores; single-sample gene set enrichment analysis) of all transcription factors with hypomethylated motifs in controls ($n = 42$) are overexpressed compared with those of 1,000 random sets of 42 transcription factors across GTEx whole-blood data ($P < 2.2 \times 10^{-16}$, Wilcoxon's Rank Sum test, two-sided). **c**, Expression profile of all transcription factors ($n = 52$) characterized as binding in patients with PDAC. Several pancreas-specific or pancreatic-cancer-associated transcription factors were identified.

Moreover, hallmark transcription factors that drive molecular subtypes of pancreatic cancer were also identified. **d**, Expression profile (ssGSEA scores) of all transcription factors with hypomethylated motifs in cases ($n = 52$) are overexpressed compared with those of 1,000 random sets of 52 transcription factors in the normal pancreas (GTEx data) (Wilcoxon Rank Sum test, two-sided test, $P < 2.2 \times 10^{-16}$). **e**, Expression profile of all transcription factors with hypomethylated motifs in PDAC cases ($n = 52$) are overexpressed compared those of 1,000 random sets of 52 transcription factors in PDAC tissue (TCGA data) (Wilcoxon Rank Sum test, two-sided test, $P < 2.2 \times 10^{-16}$). For violin plots (**b**, **d**, **e**) the ends of the boxes represent the lower and upper quartiles and the middle line indicates the median. Whiskers represent $1.5 \times$ IQR, and outliers are excluded. Rotated kernel densities are also displayed.

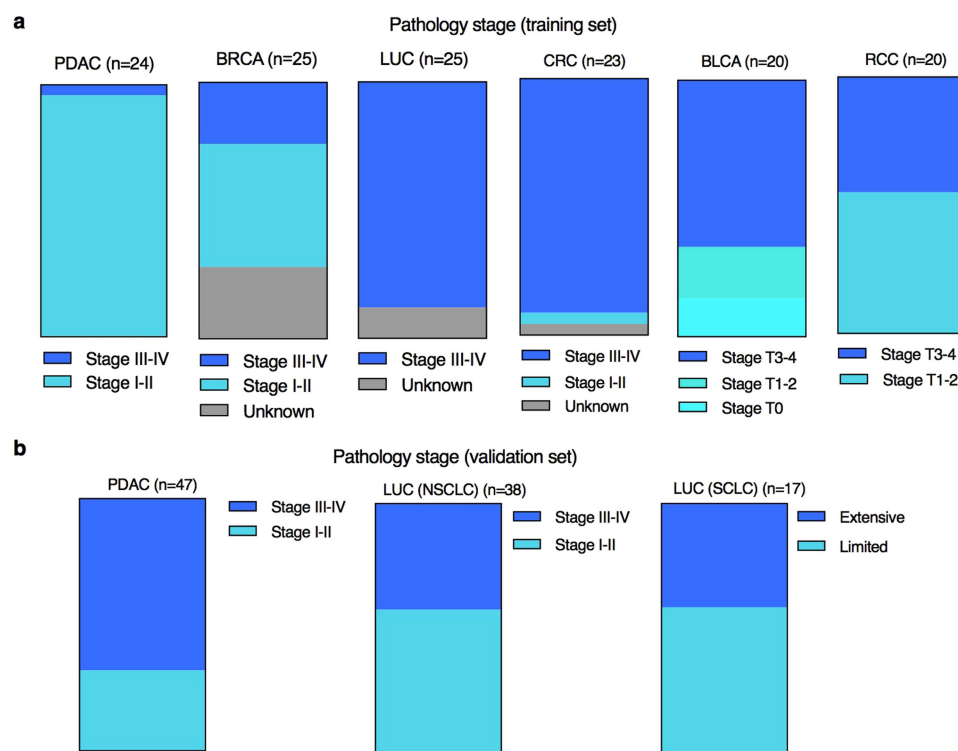


Extended Data Fig. 7 | Quality control of cfMeDIP-seq from circulating cfDNA from multiple cancer types. a, c, e, g, i, k, Specificity of the reaction; and b, d, f, h, j, CpG enrichment score for each sample per cancer type. The horizontal dashed lines represent a CpG enrichment of 1.



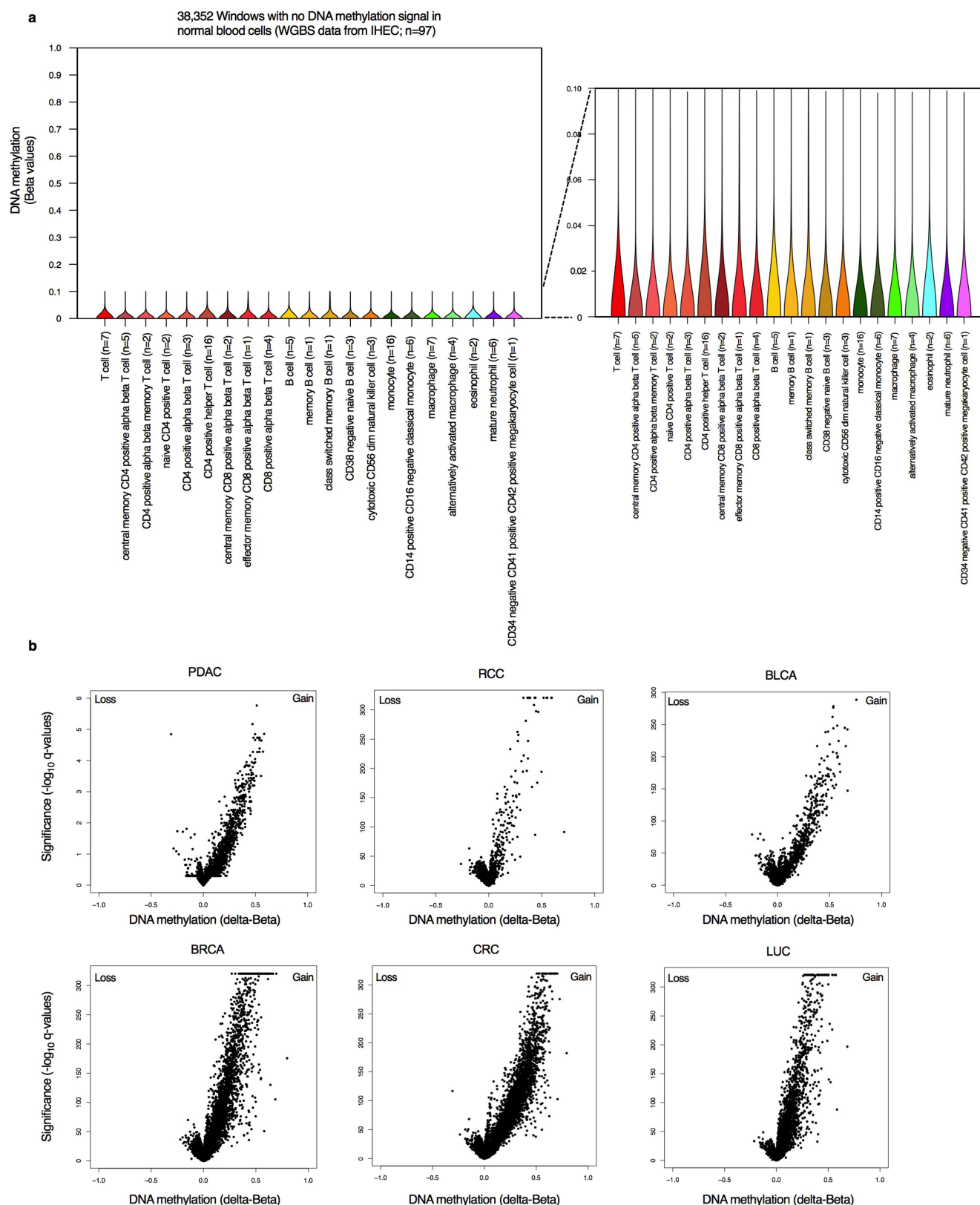
Extended Data Fig. 8 | Comparison of plasma cfDNA DMRs with tumour DMCs. **a**, Yield of cfDNA extracted per ml of plasma from healthy donors ($n = 24$), bladder cancer ($n = 20$), renal cancer ($n = 20$), lung cancer ($n = 25$), breast cancer ($n = 25$), pancreatic cancer ($n = 24$), colorectal cancer (23) and AML ($n = 28$). Horizontal bars represent the mean, with dots representing individual samples. **b–h**, Scatter plots showing the DNA methylation difference for all overlapping windows in PDAC ($n = 245,980$ windows) (**b**), AML ($n = 206,735$ windows) (**c**),

BLCA ($n = 193,943$ windows) (**d**), BRCA ($n = 204,623$ windows) (**e**), CRC ($n = 210,645$ windows) (**f**), LUC ($n = 193,043$ windows) (**g**) and RCC ($n = 198,390$ windows) (**h**). The x axis shows the DNA methylation difference between the primary tumour (TCGA data) and normal PBMCs. The y axis shows the DNA methylation difference between the plasma cfDNA methylation for each cancer type and healthy controls from the cfMeDIP-seq data. The blue line is a trend line, with the correlation determined by Pearson's correlation.



Extended Data Fig. 9 | Circulating plasma cfDNA methylation samples used to distinguish between multiple cancer types and healthy donors. a, b, Pathology stage (according to the AJCC/UICC 7th Edition)

breakdown by tumour type for samples in the training set (a) and in the validation set (b). Non-small-cell lung carcinoma, LUC (NSCLC); small-cell lung cancer, LUC (SCLC).



Extended Data Fig. 10 | Characterization of hypermethylated regions from cfDNA that are not methylated in leukocytes. a, Violin plots for the DNA methylation (plotted as beta value) of 38,352 regions in normal blood cells selected on the basis of low DNA methylation levels using IHEC whole-genome bisulfite sequencing data. For violin plots, the ends of the boxes represent the lower and upper quartiles and the middle line represents the median. Whiskers represent $1.5 \times$ IQR, and outliers are excluded. Rotated kernel densities are also displayed. **b**, Volcano plots representing the regions with low DNA methylation levels in normal blood

cells that overlap with hypermethylated regions in the plasma cfDNA for PDAC ($n = 3,146$ CpG sites) relative to normal tissue, and RCC ($n = 2,767$ CpG sites), BLCA ($n = 3,286$ CpG sites), BRCA ($n = 6,836$ CpG sites), CRC ($n = 8,360$ CpG sites) and LUC ($n = 5,239$ CpG sites) relative to PBMCs. The x axis represents DNA methylation (plotted as delta beta value), obtained from tumour data from TCGA for cancers other than PDAC and RRBS for PDAC. The y axis represents $-\log_{10} q$ values (Benjamini Hochberg false discovery rate, BH-FDR).

The entropic force generated by intrinsically disordered segments tunes protein function

Nicholas D. Keul¹, Krishnadev Oruganty², Elizabeth T. Schaper Bergman³, Nathaniel R. Beattie¹, Weston E. McDonald¹, Renuka Kadirvelraj¹, Michael L. Gross³, Robert S. Phillips⁴, Stephen C. Harvey⁵ & Zachary A. Wood^{1*}

Protein structures are dynamic and can explore a large conformational landscape^{1,2}. Only some of these structural substates are important for protein function (such as ligand binding, catalysis and regulation)^{3–5}. How evolution shapes the structural ensemble to optimize a specific function is poorly understood^{3,4}. One of the constraints on the evolution of proteins is the stability of the folded ‘native’ state. Despite this, 44% of the human proteome contains intrinsically disordered peptide segments greater than 30 residues in length⁶, the majority of which have no known function^{7–9}. Here we show that the entropic force produced by an intrinsically disordered carboxy terminus (ID-tail) shifts the conformational ensemble of human UDP- α -D-glucose-6-dehydrogenase (UGDH) towards a substate with a high affinity for an allosteric inhibitor. The function of the ID-tail does not depend on its sequence or chemical composition. Instead, the affinity enhancement can be accurately predicted based on the length of the intrinsically disordered segment, and is consistent with the entropic force generated by an unstructured peptide attached to the protein surface^{10–13}. Our data show that the unfolded state of the ID-tail rectifies the dynamics and structure of UGDH to favour inhibitor binding. Because this entropic rectifier does not have any sequence or structural constraints, it is an easily acquired adaptation. This model implies that evolution selects for disordered segments to tune the energy landscape of proteins, which may explain the persistence of intrinsic disorder in the proteome.

Intrinsically disordered segments can exhibit complex functions such as ligand binding, scaffolding of multi-protein complexes and mediating allosteric regulation^{14–18}. However, many intrinsically disordered segments are assumed to be nonfunctional and are often removed from proteins to facilitate structural studies. For example, the 30-residue disordered C terminus of UGDH (residues 465–494) is often removed with no apparent impact on kinetic parameters¹⁹. Here, we show that this C-terminal segment (called the ID-tail) plays a role in the allosteric mechanism of UGDH. UGDH catalyses the NAD⁺-dependent oxidation of UDP- α -D-glucose (UDP-Glc) to UDP- α -D-glucuronic acid¹⁹, and is regulated by the allosteric feedback inhibitor UDP- α -D-xylose (UDP-Xyl)^{20,21}. Three UGDH dimers associate to form an inactive hexamer (E*)^{22–26} (Fig. 1a, b). The binding of substrate induces an allosteric switch (T131-loop– α 6 helix) in the E* hexamer to produce the active state (E)^{22,23,26,27} (Fig. 1a, c). The allosteric inhibitor UDP-Xyl competes with UDP-Glc for the active sites, and upon binding, triggers the allosteric switch to produce the inhibited state (E ^{Ω})^{22,24,25,27}. This inhibition mechanism is atypical in that the active site also functions as an allosteric site to control the structure and activity of the hexamer^{22–27} (Fig. 1a, c). The E ^{Ω} state has a high affinity for UDP-Xyl and a low affinity for UDP-Glc^{22,27}. Therefore, the allosteric transition of the inhibited E ^{Ω} hexamer to the E state can be observed as cooperativity in substrate saturation curves^{22,27}. We compared the structure and activity of full-length UGDH (UGDH(FL)) to a construct lacking the ID-tail (UGDH(Δ ID)). We solved the structures of E* states of

UGDH(FL) and UGDH(Δ ID) in isomorphous crystal lattices, showing that there were no substantial differences (Extended Data Fig. 1a–c). UGDH(FL) and UGDH(Δ ID) also have a similar catalytic rate constant (k_{cat}) and Michaelis constant (K_m) for both substrate and coenzyme, consistent with earlier reports¹⁹ (Extended Data Table 2). By contrast, the allosteric response is sensitive to the ID-tail; deletion of the ID-tail reduces the affinity for UDP-Xyl by an order of magnitude (Fig. 1d). UGDH(Δ ID) still binds UDP-Glc cooperatively, indicating that the deletion of the ID-tail reduces UDP-Xyl affinity but does not prevent the formation of the E ^{Ω} hexamer (Fig. 1d, Extended Data Fig. 2a, b).

Both the ID-tail and the α 6 helix of the allosteric switch are located in the hexamer-building interface between adjacent dimers, suggesting that these two elements may work together to increase the affinity for UDP-Xyl (Fig. 1b). We used the allosteric-quenching A136M substitution to determine whether the ID-tail functions independently of the allosteric switch. This substitution has been shown to lock the allosteric switch and the hexamer in the low UDP-Xyl-affinity E state²². Inhibition studies show no marked difference between the UDP-Xyl affinities of UGDH(FL/A136M) and UGDH(Δ ID/A136M), which suggests that the ID-tail requires both a functional allosteric switch and the E ^{Ω} state to enhance the affinity for UDP-Xyl (Fig. 1d).

The location of the α 6 helix in the hexamer-building interface suggests that the oligomeric structure might be important for the function of the ID-tail (Fig. 1b). Sedimentation velocity studies show that the UGDH(Δ ID) E* hexamer is slightly less stable than the UGDH(FL) E* hexamer, perhaps explaining its reduced affinity for UDP-Xyl (Extended Data Fig. 3a). We tested the role of the hexamer with the M11 interfacial loop substitution, which prevents hexamer formation and stabilizes the dimer (UGDH(FL-dimer) and UGDH(Δ ID-dimer))²⁷. UDP-Xyl binds to the UGDH(FL-dimer) with sevenfold higher affinity than UGDH(Δ ID-dimer), demonstrating that the ID-tail does not require the hexamer to enhance the affinity for UDP-Xyl (Fig. 1d and Supplementary Information Section 1).

The ID-tail is highly conserved in vertebrate UGDHs (Fig. 2a). We examined the importance of primary structure in the ID-tail by randomizing the native sequence to create two distinct ID-tails (UGDH(R1) and UGDH(R2)) (Fig. 2b). Surprisingly, the UGDH(FL), UGDH(R1) and UGDH(R2) constructs have similar affinities for UDP-Xyl (Fig. 2c). Next, all six prolines in the ID-tail were substituted with serine (UGDH(–Pro)) (Fig. 2b). Because serine and proline both promote disorder^{28,29}, this substitution conserves the unfolded state and disrupts any possible proline-specific interactions. Analysis of UGDH(–Pro) shows that the prolines do not contribute to UDP-Xyl affinity (Fig. 2c). Because all of the above constructs conserve the positive charge of the native ID-tail ($pI = 10.1$), we created a negatively charged ID-tail ($pI = 4.4$) using lysine-to-serine substitutions (UGDH(–Lys)) (Fig. 2b). Despite the charge switch, there was still no substantial change in UDP-Xyl affinity (Fig. 2c). Finally, we replaced the ID-tail with polyserine (UGDH(Ser)) without causing a marked change in UDP-Xyl affinity (Fig. 2b, c). Therefore, the conserved primary structure is not required

¹Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA. ²Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA. ³Department of Chemistry, Washington University in St. Louis, St. Louis, MO, USA. ⁴Department of Chemistry, University of Georgia, Athens, GA, USA. ⁵Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA. *e-mail: zaw@uga.edu

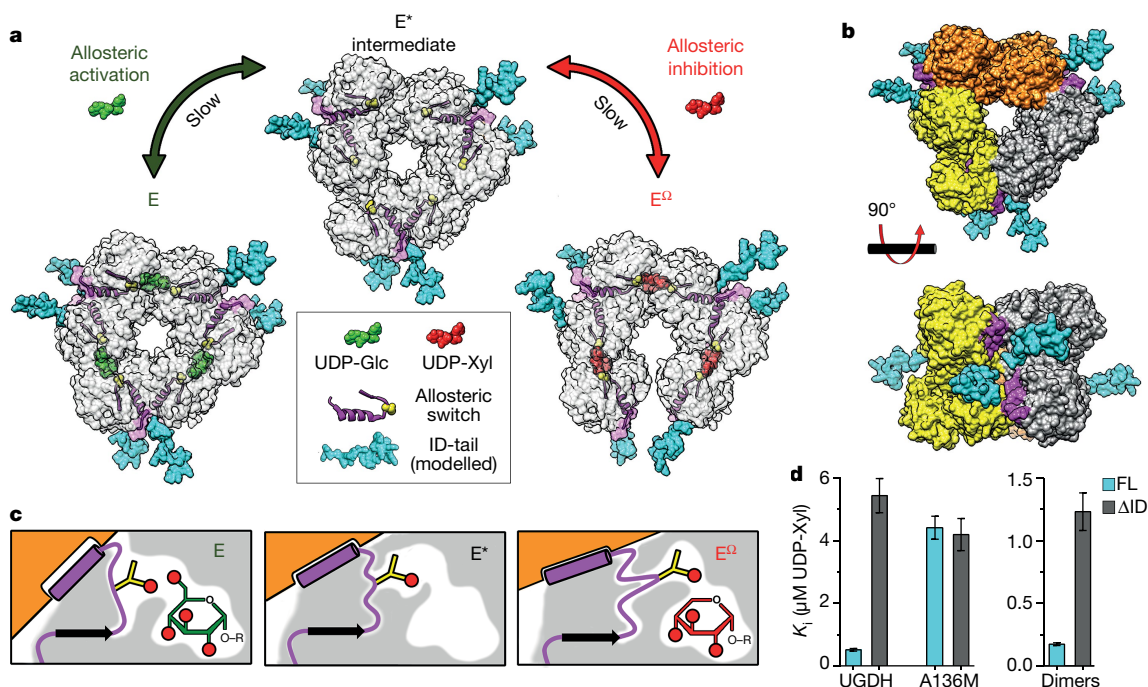


Fig. 1 | The role of the ID-tail in allosteric inhibition of UGDH.

a, Unliganded UGDH forms an inactive (E^*) hexamer. UDP-Glc (green) induces the Thr131-loop- $\alpha 6$ allosteric switch (yellow spheres and magenta ribbons and surface) to slowly isomerize into the active (E) state. UDP-Xyl (red) competes with UDP-Glc for the active site, and induces the allosteric switch to slowly isomerize into the inhibited (E^Δ) state. The slow isomerizations are due to the repacking of the allosteric switch in the protein core^{22,24,25,27}. Because the ID-tail is disordered in the E , E^* and E^Δ states (Extended Data Fig. 1 and refs^{22,24–26}), we have modelled energy-minimized conformations of the ID-tail (cyan) onto the structures of UGDH to depict the proximity to the active site, hexamer-building interface and the allosteric switch. **b**, Top and side view of the UGDH E^* hexamer that forms from the association of three dimers (orange, grey

and yellow)^{22–27}. The ID-tail of each dimer is located near two allosteric switches in the hexamer-building interface. **c**, The allosteric switch (magenta) is buried in the protein core (grey shading), which changes conformation in the E , E^* and E^Δ states. Thr131 (yellow sticks) responds to the presence or absence of the C6' OH in UDP-Glc (green) or UDP-Xyl (red), respectively. This response shifts the $\alpha 6$ helix (magenta cylinder) in the hexamer-building interface, which rotates the adjacent subunit (orange) to produce the E or E^Δ hexamer, as appropriate. Red circles depict hydroxyl (OH) groups. **d**, The UDP-Xyl affinity depends on the ID-tail and allostery. Data are the globally fit $K_i \pm \text{s.e.m.}$ derived from two or three independent rate curves with varying amounts of inhibitor ($n \geq 31$ independent data points; see Extended Data Table 2 for specific values).

for UDP-Xyl affinity, but may have been selected for because of an additional, unrelated function in vivo (Fig. 2a). The absence of sequence constraints argues against any mechanism in which the ID-tail specifically interacts with the inhibitor or the protein.

Next, we considered the possibility that the ID-tail might enhance UDP-Xyl affinity through a sequence-independent interaction

involving the polypeptide backbone. Because the six prolines in the UGDH(FL), UGDH(R1) and UGDH(R2) ID-tails sample 16 unique positions throughout the sequence without altering UDP-Xyl affinity, it is unlikely that a backbone-specific interaction is important for function (Fig. 2b, c). Nevertheless, if there was a backbone-specific interaction, a plot of affinity versus ID-tail length would reveal a discontinuity

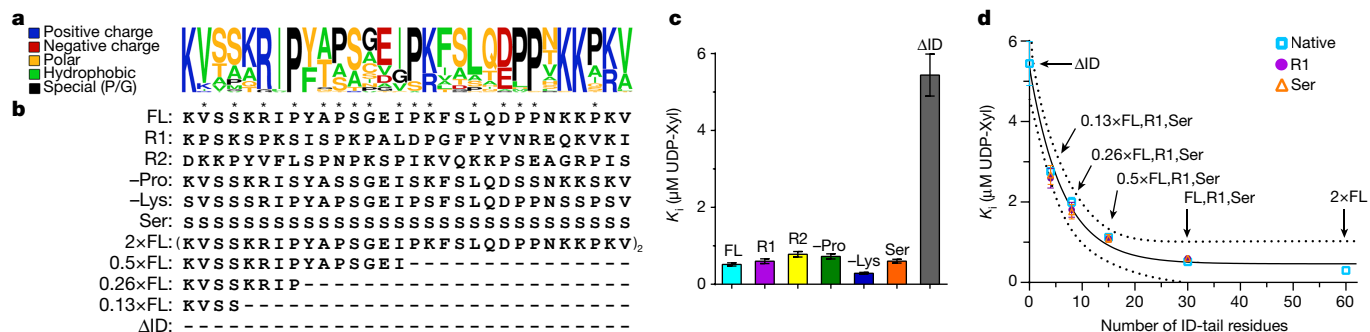


Fig. 2 | Structural constraints of the ID-tail. **a**, Alignments of the 30-residue ID-tail sequence (residues 465–494) from 79 vertebrate UGDHs (Extended Data Fig. 4a, b). Residues are coloured according to type, and the height of each residue represents the relative frequency. Alignments were generated using the WebLogo server (<http://weblogo.berkeley.edu>). **b**, Sequence modifications made to the primary structure of the ID-tail (Extended Data Fig. 4b). Asterisks indicate positions in the sequence that are sampled with a proline residue in either UGDH(FL), UGDH(R1) or UGDH(R2). **c**, UDP-Xyl affinity is independent of the ID-tail sequence. Data are the globally fit $K_i \pm \text{s.e.m.}$ derived from two

or three independent rate curves with varying amounts of inhibitor. See Extended Data Table 2 for the specific number of independent data points ($n \geq 27$). **d**, The affinity for UDP-Xyl depends on the length of the ID-tail. Data are the globally fit $K_i \pm \text{s.e.m.}$ derived from three independent rate curves with varying amounts of inhibitor ($n \geq 38$ independent data points; see Extended Data Table 2 for specific values). For some points the s.e.m. is smaller than the data label. The data were fit to equation (3) (solid line) with 95% confidence intervals indicated (dashed lines). The fit predicts a maximum affinity of $0.46 \pm 0.18 \mu\text{M}$, corresponding to a free-energy change of $-1.45 \text{ kcal mol}^{-1}$.

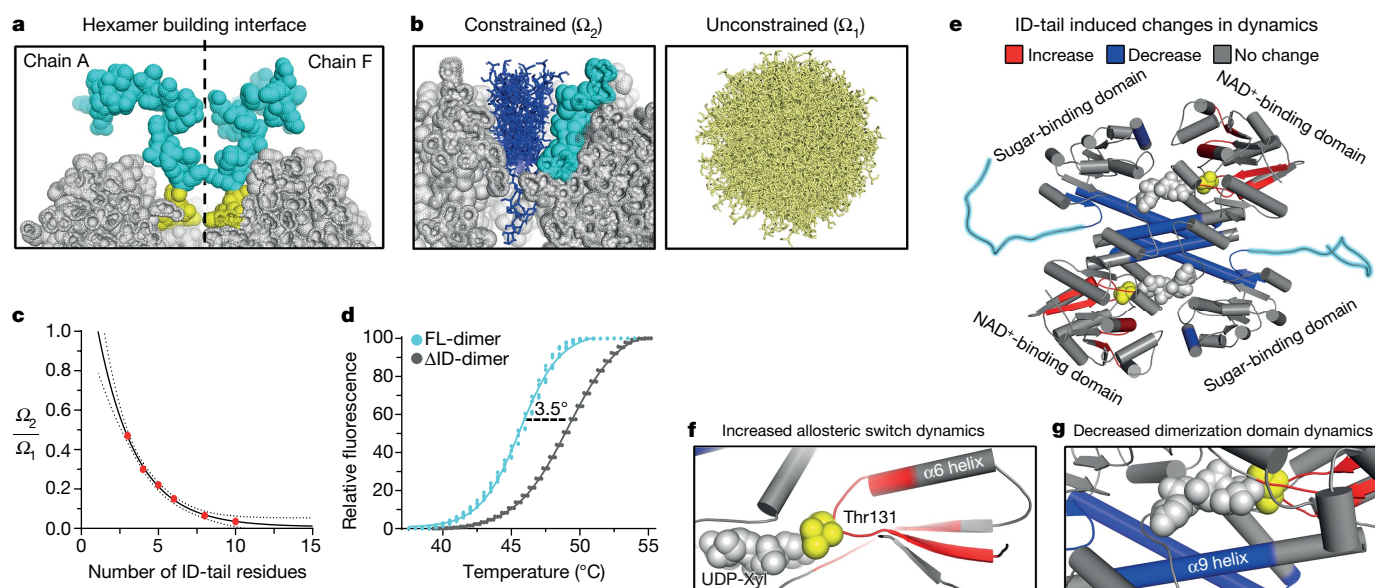


Fig. 3 | The entropic force of the ID-tail alters the structure of UGDH. **a**, Cut-away of the UGDH surface (grey spheres) at the hexamer-building interface (dashed lines), depicting the modelled ID-tails (cyan and yellow spheres) from adjacent subunits (grey, chains A and F). The volume-exclusion effect of the hexamer-building interface tightly constrains the conformations of the first four disordered residues (465–468) of the ID-tail (yellow). **b**, Left, a representative subset of the surface-constrained conformations of a 10-residue ID-tail (blue sticks) from Monte Carlo sampling (see Methods for details). The adjacent ID-tail is shown as cyan spheres. Right, a representative sampling of accessible conformations without surface constraints (see also Extended Data Fig. 5). **c**, The fraction of constrained ID-tail conformations (Ω_2) over the possible conformations of a free ID-tail (Ω_1) exponentially converges

when the critical segment was removed. Inhibition studies comparing UGDH(FL), UGDH(Δ ID) and three new constructs with ID-tails of varying length (UGDH($2 \times$ FL), UGDH($0.5 \times$ FL), UGDH($0.26 \times$ FL) and UGDH($0.13 \times$ FL), shown in Fig. 2b) show that the affinity can be modelled as a simple exponential decay (Fig. 2d). We confirmed that this saturable effect is independent of sequence by using polyserine ID-tails of corresponding lengths (UGDH(Ser), UGDH($0.5 \times$ Ser), UGDH($0.26 \times$ Ser) and UGDH($0.13 \times$ Ser)) and similarly, using corresponding lengths of the scrambled R1 construct (UGDH(R1), UGDH($0.5 \times$ R1), UGDH($0.26 \times$ R1) and UGDH($0.13 \times$ R1)) (Fig. 2d). It is notable that UGDH($0.13 \times$ FL), UGDH($0.13 \times$ Ser) and UGDH($0.13 \times$ R1) still enhance UDP-Xyl binding affinity; the conformations of these short, four-residue ID-tails are tightly constrained within a surface pocket, which should stabilize any weak structure (Fig. 3a). Nevertheless, none of the E, E* and E^Ω UGDH(FL) crystal structures (42 unique chains) show evidence of an ordered interaction within the pocket (Extended Data Fig. 1 and refs ^{22,24–26}).

The data presented so far provide strong evidence that the high-affinity binding of UDP-Xyl is a function of the unfolded state of the ID-tail. An unstructured polymer tethered to a surface generates an entropic force at the point of attachment^{10–12}, which can be strong enough to distort lipid bilayers³⁰ and alter protein stability¹³. This force originates from the volume exclusion effects of the surface, which reduce the conformational entropy of the attached polymer (Fig. 3b). Because the entropy of the polymer increases with distance from the surface, the entropic force converges to a maximum value as the chain length increases^{10–12}. The unfavourable change in free energy produced by constraining an unstructured, non-interacting peptide ($\Delta G_{\text{constrained}}$) is

$$\Delta G_{\text{constrained}} = -RT \ln \left(\frac{\Omega_2}{\Omega_1} \right) \quad (1)$$

with increasing ID-tail length. The data were fit to an exponential decay (Extended Data Fig. 5c). **d**, The ID-tail destabilizes UGDH by 3.5 °C. **e**, Comparing HDX rates of UGDH(FL-dimer) and UGDH(Δ ID-dimer) shows that the ID-tail (cyan) alters the structure and dynamics of UGDH. Peptides displaying increases (red), decreases (blue) and no change (grey) in HDX rates are mapped to the structure. UDP-Xyl (grey spheres) was not used in the assay but is modelled in the active site. Thr131 of the allosteric switch is shown as yellow spheres. **f**, Close-up view of the allosteric switch (Thr131– $\alpha 6$ helix), which shows an increase in HDX rates. **g**, Close-up view of the dimerization domain, which is largely inaccessible to solvent. Data shown in **e–g** were derived from the normalized cumulative per cent deuterium uptake (%D) comparing UGDH(FL-dimer) and UGDH(Δ ID-dimer) (Extended Data Fig. 6).

where Ω_1 is the sum of all possible states of an unconstrained peptide and Ω_2 is the subset of states constrained by the protein surface and the adjacent ID-tail (RT is the product of the molar gas constant, R , and the temperature, T). Using Monte Carlo sampling of coarse-grained, sterically allowed bins of ϕ and ψ torsion angles we calculated the fraction of constrained conformations for various ID-tail lengths (see Methods, Fig. 3b, Extended Data Fig. 5). For this simulation, the adjacent ID-tail was held in a fixed conformation (Extended Data Fig. 5). If the conformational entropy of the ID-tail contributes to the change in UDP-Xyl affinity, then we would expect Ω_2/Ω_1 and the affinity constant K_i to display similar behaviour with increasing tail length. Despite the simplicity of the Monte Carlo model, the simulations confirm that Ω_2/Ω_1 converges as the ID-tail length increases (Fig. 3c).

Studies have shown that the entropic force generated by a tethered polymer can alter protein stability¹³. We carried out thermal denaturation studies of UGDH dimers (chosen to avoid complications arising from hexamer dissociation), and found that the high-affinity UGDH(FL-dimer) ($K_i = 0.17 \mu\text{M}$) is less stable than the low-affinity UGDH(Δ ID-dimer) ($K_i = 1.23 \mu\text{M}$) (Fig. 3d). The destabilizing effect of the ID-tail should also be reflected in the structure and dynamics of UGDH. To examine these changes at the peptide level, we compared the hydrogen–deuterium exchange (HDX) rates of UGDH(FL-dimer) and UGDH(Δ ID-dimer) using mass spectrometry. As expected, the fragment corresponding to the ID-tail is fully exchanged in less than 120 s, which is consistent with a disordered peptide³¹ (Extended Data Fig. 6a). The ID-tail increases the HDX rates of several segments in the NAD⁺ binding domain, with the largest increases occurring in the allosteric switch and an adjacent peptide (Fig. 3e–g). An increase in HDX rates for a buried peptide such as the allosteric switch and the surrounding segments indicates an increase in the overall dynamics of the domain. This is notable, because the binding of UDP-Xyl induces the allosteric switch and surrounding core residues to change conformation

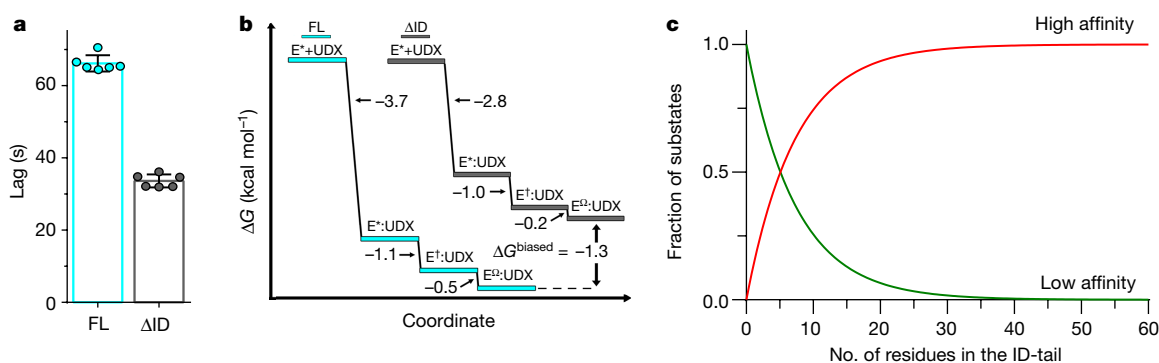


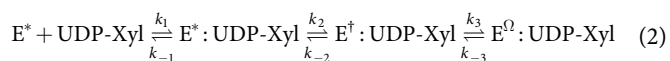
Fig. 4 | The ID-tail shapes the conformational landscape of UGDH.

a, The ID-tail increases the hysteresis of allosteric activation (E^* to E). Data are mean \pm s.d. ($n = 6$ independent experiments). **b**, Free-energy plot of the transient-state kinetic model for the allosteric inhibition with UDP-Xyl (UDX) (equation (2)) of UGDH(FL) (E^*_{FL}) and UGDH(Δ ID) ($E^*_{\Delta ID}$). The free energy for the initial binding step (second order)

using the standard state defined in units of mM, and the two isomerization steps (first order) were determined from the equilibrium constants K_1 , K_2 and K_3 defined in Extended Data Fig. 7e. **c**, Plot of the high- and low-affinity fractional components of equation (3), showing that the shift in states is a function of length of the ID-tail.

and repack into the high affinity E^Ω state^{22,27} (Fig. 1a, c). The ID-tail also decreases the HDX rates of several segments in the dimerization and sugar-binding domains, suggesting that these areas become more structured (Fig. 3e, g). The largest decrease is observed in the $\alpha 9$ helix of the dimerization domain (residues 222–240). This helix is largely inaccessible to solvent in crystal structures, which suggests that the ID-tail reduces the overall dynamics of the dimer interface (Fig. 3g). Overall, the data show that the cost of constraining the ID-tail destabilizes a low-affinity substate, which biases the conformational ensemble towards a structurally and dynamically distinct high-affinity substate. A simple exponential fit of Ω_2/Ω_1 (Fig. 3c) shows that the energetic cost of constraining the ID-tail converges to approximately 2.4 kcal mol^{−1} (equation (1)). Therefore, our simple Monte Carlo model supports the argument that entropic confinement effects generate sufficiently strong forces to explain the maximum expected gain in UDP-Xyl binding affinity of -1.45 kcal mol^{−1} (Figs. 2d, 3c, Extended Data Fig. 5). More rigorous calculations on other systems using simpler polymer models (and simpler confinement geometries) also find confinement free-energy costs of the same magnitude^{32,33}.

If the ID-tail favours the dynamics associated with the repacking of the allosteric switch into the E^Ω state, then we would expect to see a difference in the activation (E^* to E) and inhibition kinetics (E^* to E^Ω) (Fig. 1a). Pre-steady-state analysis of progress curves shows that the ID-tail slows the rate of activation hysteresis (E^* to E) by 39% (Fig. 4a). Next, we examined the UDP-Xyl-induced isomerization of UGDH to the E^Ω state. Transient-state analysis of UDP-Xyl binding kinetics revealed a three-phase exponential decay of UGDH time-resolved tryptophan fluorescence, and the data were globally fit by computer simulation (see Methods and Extended Data Fig. 7a–e). The same kinetic model produced the best fit for both UGDH(FL) and UGDH(Δ ID) and predicts UDP-Xyl affinities that are consistent with our steady-state inhibition studies (Extended Data Fig. 7):



Where k_n is the rate constant for reaction n . According to this model, UDP-Xyl binds to the E^* state and induces two sequential isomerizations. On the basis of the allosteric model, we had expected a single isomerization from E^* to the E^Ω state (Fig. 1a). We call the additional transient E^\dagger ; it represents an intermediate between the E^* and E^Ω states. The ID-tail changes the kinetic parameters of each transient observed in the time-resolved fluorescence (Extended Data Fig. 7e). The largest effect of the ID-tail is a 4.4-fold enhancement of the initial UDP-Xyl binding step, corresponding to a -0.9 kcal mol^{−1} gain in affinity (Fig. 4b). The kinetic model predicts an overall favourable gain in binding affinity of -1.3 kcal mol^{−1}, which agrees well with the observed

gain of -1.39 kcal mol^{−1} (Fig. 4b, Extended Data Table 2). The different stabilities of the corresponding UGDH(FL) and UGDH(Δ ID) transients, combined with the fact that the ID-tail slows activation hysteresis and accelerates inhibition kinetics, supports our conclusion that the ID-tail alters the energy landscape to favour inhibition by UDP-Xyl (Fig. 4c).

Collectively, our data supports a model in which the entropic force of the ID-tail rectifies the energy landscape of UGDH to favour a substate with a high affinity for UDP-Xyl. We can now interpret the exponential curve in Fig. 2d as follows:

$$K_i(l) = K_i^{\text{unbiased}} e^{-kl} + K_i^{\text{biased}} (1 - e^{-kl}) \quad (3)$$

This implies that: (i) UGDH exists as an ensemble of low-affinity (K_i^{unbiased}) and high-affinity (K_i^{biased}) substates; (ii) the ID-tail functions as a length (l)-dependent entropic rectifier that shifts (with bias k) the distribution towards the high affinity substate; and (iii) the observed UDP-Xyl affinity results from a fractional summation of the low and high affinity substates at a given ID-tail length (Fig. 4c). The fit to equation (3) produces a K_i^{biased} of 0.46 ± 0.18 μ M UDP-Xyl, which corresponds to a maximum favourable gain in binding energy of approximately -1.45 kcal mol^{−1}. The lack of sequence constraints implies that the entropic force of any intrinsically disordered segment is capable of shaping the conformational ensemble of a protein. In fact, an N-terminal hexahistidine affinity tag has been shown to alter the internal dynamics of a myoglobin³⁴. Thus, the persistence of low-complexity intrinsically disordered segments in the proteome may reflect the selection for entropic rectifiers that can tune the function of a protein by shaping the native-state ensemble.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-018-0699-5>.

Received: 24 August 2017; Accepted: 10 September 2018;
Published online 12 November 2018.

1. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
2. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
3. Campbell, E. et al. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944–950 (2016).
4. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
5. Kumar, S., Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **9**, 10–19 (2000).
6. Oates, M. E. et al. D²P²: database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–D516 (2013).

7. van der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
8. Papaleo, E. et al. The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* **116**, 6391–6423 (2016).
9. He, B. et al. Predicting intrinsic disorder in proteins: an overview. *Cell Res.* **19**, 929–949 (2009).
10. Bickel, T., Jeppesen, C. & Marques, C. M. Local entropic effects of polymers grafted to soft interfaces. *Eur. Phys. J. E* **4**, 33–43 (2001).
11. Bickel, T., Marques, C. & Jeppesen, C. Pressure patches for membranes: the induced pinch of a grafted polymer. *Phys. Rev. E* **62**, 1124–1127 (2000).
12. Waters, J. T. & Kim, H. D. Calculation of a fluctuating entropic force by phase space sampling. *Phys. Rev. E* **92**, 013308 (2015).
13. Carmichael, S. P. & Shell, M. S. Entropic (de)stabilization of surface-bound peptides conjugated with polymers. *J. Chem. Phys.* **143**, 243103 (2015).
14. Ferreon, A. C., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390–394 (2013).
15. Hilser, V. J. An ensemble view of allostery. *Science* **327**, 653–654 (2010).
16. Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl Acad. Sci. USA* **104**, 8311–8315 (2007).
17. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).
18. Li, J. et al. Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. *eLife* **6**, e30688 (2017).
19. Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U. & Nidetzky, B. Structure and mechanism of human UDP-glucose 6-dehydrogenase. *J. Biol. Chem.* **286**, 23877–23887 (2011).
20. Gainey, P. A. & Phelps, C. F. Interactions of uridine-diphosphate glucose dehydrogenase with inhibitor uridine-diphosphate xylose. *Biochem. J.* **145**, 129–134 (1975).
21. Neufeld, E. F. & Hall, C. W. Inhibition of UDP-D-glucose dehydrogenase by UDP-D-xylose - a possible regulatory mechanism. *Biochem. Biophys. Res. Commun.* **19**, 456–461 (1965).
22. Beattie, N. R., Keul, N. D., Sidlo, A. M. & Wood, Z. A. Allostery and hysteresis are coupled in human UDP-glucose dehydrogenase. *Biochemistry* **56**, 202–211 (2017).
23. Kadirvelraj, R., Sennett, N. C., Custer, G. S., Phillips, R. S. & Wood, Z. A. Hysteresis and negative cooperativity in human UDP-glucose dehydrogenase. *Biochemistry* **52**, 1456–1465 (2013).
24. Kadirvelraj, R., Sennett, N. C., Polizzi, S. J., Weitzel, S. & Wood, Z. A. Role of packing defects in the evolution of allostery and induced fit in human UDP-glucose dehydrogenase. *Biochemistry* **50**, 5780–5789 (2011).
25. Sennett, N. C., Kadirvelraj, R. & Wood, Z. A. Conformational flexibility in the allosteric regulation of human UDP- α -D-glucose 6-dehydrogenase. *Biochemistry* **50**, 9651–9663 (2011).
26. Sennett, N. C., Kadirvelraj, R. & Wood, Z. A. Cofactor binding triggers a molecular switch to allosterically activate human UDP- α -D-glucose 6-dehydrogenase. *Biochemistry* **51**, 9364–9374 (2012).
27. Kadirvelraj, R. et al. Hysteresis in human UDP-glucose dehydrogenase is due to a restrained hexameric structure that favors feedback inhibition. *Biochemistry* **53**, 8043–8051 (2014).
28. Uversky, V. N. The intrinsic disorder alphabet. III. Dual personality of serine. *Intrinsically Disord. Proteins* **3**, e1027032 (2015).
29. Theillet, F. X. et al. The alphabet of intrinsic disorder: I. Act like a pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord. Proteins* **1**, e24360 (2013).
30. Busch, D. J. et al. Intrinsically disordered proteins drive membrane curvature. *Nat. Commun.* **6**, 7875 (2015).
31. Balasubramaniam, D. & Komives, E. A. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta* **1834**, 1202–1209 (2013).
32. Chen, J. Z. Y. Theory of wormlike polymer chains in confinement. *Prog. Polym. Sci.* **54–55**, 3–46 (2016).
33. Smyda, M. R. & Harvey, S. C. The entropic cost of polymer confinement. *J. Phys. Chem. B* **116**, 10928–10934 (2012).
34. Thielges, M. C., Chung, J. K., Axup, J. Y. & Fayer, M. D. Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry* **50**, 5799–5805 (2011).

Acknowledgements The authors thank A. P. Karplus, B. W. Matthews, S. N. Savvides and the members of the Z.A.W. laboratory for helpful discussions. We also thank R. Wang of Norclone for producing the R1 truncation constructs. This work was supported by the NIH National Institute of General Medicine grants R01GM114298 awarded to Z.A.W. and P41GM103422 awarded to M.L.G.

Reviewer information *Nature* thanks J. Gsponer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions N.D.K. and Z.A.W. designed the study, analysed the data and composed the manuscript. N.D.K. performed the majority of the experiments. K.O. performed all the bioinformatic analyses and conducted the simulations of the ID-tail and analysed the results with S.C.H. S.C.H. contributed to the development of the entropic force model. E.T.S.B. performed HDX mass spectrometry experiments and interpreted the results with M.L.G. Both offered direction in HDX experiment design and contributed to this portion of the manuscript. M.L.G. further contributed with edits to the larger work. N.R.B. performed analysis and refinement of crystal structures. W.E.M. conducted several AUC experiments and performed thermal denaturation assays. R.K. solved Protein Data Bank entry 5VR8. R.S.P. contributed to the development of our kinetic model.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0699-5>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0699-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to Z.A.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression, purification, and quantification of UGDH constructs.

All UGDH coding sequences were synthesized and cloned into pET-15b vectors (Nordlone). Sequences contained an N-terminal hexahistidine affinity tag adjacent to a tobacco etch virus (TEV) cleavage site. The expression and purification of UGDH constructs were conducted under identical conditions as previously described^{22–27}. Following purification, the N-terminal hexahistidine tag was cleaved with TEV protease. An additional immobilized metal affinity column (IMAC) was used to obtain the pure, His-tag-free protein. Unless otherwise noted, all proteins were dialysed into a storage buffer (25 mM Tris pH 8.0 and 50 mM NaCl) and concentrated to ≥ 20 mg/ml. Proteins were quantified in dilution replicates ($n \geq 6$) using their respective molar extinction coefficients, based on their specific amino acid composition³⁵.

Protein crystallization, data collection, and structure solution. To crystallize the E* conformation of UGDH (Δ ID), the protein (10.4 mg/ml) was dialysed into 20 mM MES pH 5.6, 150 mM NaCl and crystallized at 20 °C using free interface diffusion in a 1.0-mm capillary containing 5 μ l of 10.4 mg/ml enzyme and 200 μ l of precipitant solution (100 mM MES pH 6.2, 100 mM MgCl₂, and 16% PEG 3350). Crystals were cryoprotected in the precipitant solution supplemented with 18% glycerol and then plunged into liquid nitrogen. A 2.64 Å resolution dataset was collected on the 22-ID beamline (SER-CAT) at the Argonne National Laboratory using an MAR 300-mm CCD detector. The data were processed in space group C2 using XDS³⁶, and 5% of the data were set aside for cross-validation³⁷. The crystal parameters and data collection statistics are summarized in Extended Data Table 1. The structure was solved by molecular replacement using the PHENIX software suite³⁸ and human UGDH (Protein Data Bank (PDB) entry: 3TF5) as a search model. The structure was then subjected to iterative cycles of manual rebuilding using Coot³⁹ and automated refinement using PHENIX with both NCS restraints^{38,40}. B-factors were refined using TLS as implemented in PHENIX. Refinement statistics^{41,42} are summarized in Extended Data Table 1.

The E^Q UGDH(FL) was crystallized in the presence of 5 mM UDP-Xyl and 10 mM adenosine diphosphate at 25 °C using the hanging drop vapour diffusion method. One microlitre of protein was mixed in a 1:1 ratio with reservoir solution (0.1 M HEPES pH 7.2, 14% 1,6-hexanediol, and 10% PEG 3350). Crystals were cryoprotected in the precipitant solution supplemented with 20% glycerol and then plunged into liquid nitrogen. A 2.0 Å resolution dataset was collected on the 21-ID beamline (SER-CAT) at the Argonne National Laboratory using a MAR 300-mm CCD detector. The dataset was processed using XDS³⁶ and 5% of the data were set aside for cross validation³⁷. The data collection statistics are listed in Extended Data Table 1. The E^Q UGDH(FL) structure was solved by molecular replacement using the PDB entry 2Q3E as a search model in PHENIX³⁸, and refined as described above. Refinement statistics^{41,42} are summarized in Extended Data Table 1.

Steady-state kinetics. All steady-state kinetic assays were conducted as previously described^{22–27}. In brief, assays contained either 100 nM UGDH (FL, FL-A136M, Δ ID, Δ ID-A136M, R1, R2, –Pro, –Lys, 0.13 \times FL, 0.26 \times FL, 0.5 \times FL, 2 \times FL, 0.13 \times Ser, 0.26 \times Ser, 0.5 \times Ser, Ser, 0.13 \times R1, 0.26 \times R1 or 0.5 \times R1) or 500 nM UGDH (FL-dimer, Δ ID-dimer) in a standard reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA) with either saturating amounts of NAD⁺ or UDP-Glc (Sigma). Substrate and enzyme were incubated separately at 25 °C for 5 min, and then reactions were initiated by rapid mixing of both solutions. Progress curves were obtained by continuously monitoring NADH production at 340 nm (molar absorptivity coefficient of 6,220 M^{–1} cm^{–1}) on an Agilent 8453 UV/Vis spectrometer equipped with a Peltier temperature controller (25 °C). UGDH progress curves display hysteresis, thus the observed initial velocity (v_i) represents a transient and does not satisfy steady-state conditions. To obtain steady-state initial velocities (v_{ss}), progress curves before the depletion of 10% substrate were fit to Frieden's equation⁴³ as in previous studies^{22,27,43}:

$$P(t) = v_{ss}t - \tau(v_{ss} - v_i)\left(1 - e^{-\frac{t}{\tau}}\right) \quad (4)$$

where P is the product produced at time t , τ is the relaxation time of the lag, and the length of the lag is $e\tau$. The v_{ss} was used for determination of UGDH steady-state kinetic parameters. Data were fit using nonlinear regression analysis in PRISM (GraphPad Software).

Because the UGDH(FL-A136M), UGDH(Δ ID-A136M), UGDH(FL-dimer) and UGDH(Δ ID-dimer) constructs do not exhibit hysteresis, the observed initial velocity was used for the determination of steady-state parameters as previously described²². UDP-Glc substrate saturation curves were fit to equation (5).

$$v_o = \frac{k_{cat}[E_t][S]}{K_M + [S]} \quad (5)$$

where v_o is the initial steady state velocity (v_{ss} in equation (4)), E_t and S are the enzyme and substrate concentrations, respectively. As previously reported^{22,23,27}, the NAD⁺ saturation curves of the UGDH hexameric enzyme display negative cooperativity and were fit to the sigmoidal rate equation (equation (6)):

$$v_o = \frac{k_{cat}[E_t][S]^h}{(K_{0.5})^h + [S]^h} \quad (6)$$

where $K_{0.5}$ is the half saturation point and h represents the Hill coefficient. The determination of the K_i for the allosteric inhibitor UDP-Xyl has been previously described^{22,27}. In brief, data were globally fit to the model for competitive inhibition with cooperativity (equation (7)) using PRISM.

$$v_o = \frac{k_{cat}[E_t][S]^h}{(K_M^{app})^h + [S]^h} \text{ where } K_M^{app} = K_M \left(1 + \frac{[I]}{K_i}\right) \quad (7)$$

K_M , k_{cat} , and K_i were shared parameters in global fitting, whereas h was fit locally to each curve. The UGDH dimers (UGDH(FL-dimer) and UGDH(Δ ID-dimer)) exhibited mixed inhibition with respect to both UDP-Glc and NAD⁺, and were globally fit to equation (8).

$$v_o = \frac{(k_{cat}^{app}[E_t][S])}{(K_M^{app}) + [S]} \text{ where } k_{cat}^{app} = \frac{k_{cat}}{\left(1 + \frac{[I]}{\alpha K_i}\right)} \text{ and } K_M^{app} = K_M \left(1 + \frac{[I]}{\alpha K_i}\right) \quad (8)$$

Here, K_i refers the competitive inhibition component, and αK_i gives the noncompetitive contribution. K_M , k_{cat} , α and K_i were shared parameters for global fitting. **Sedimentation velocity.** Sedimentation velocity analysis was conducted as previously described^{22–27}. In brief, UGDH constructs were dialysed for >12 h at 4 °C into 25 mM HEPES pH 7.5 and 150 mM KCl and diluted to a final concentration of 9 μ M. In ligand-bound studies, UGDH constructs were dialysed with comparable amounts of either substrate (UDP-Glc) or allosteric inhibitor (UDP-Xyl) for >24 h. Samples were loaded into cells equipped with 12-mm double-sector Epon centrepieces and quartz windows. The cells were then loaded into an An60 Ti rotor and equilibrated to 20 °C for 1 h. Sedimentation velocity data were collected at 50,000 r.p.m. in an Optima XLA analytical ultracentrifuge for 8–12 h. Data were recorded at 280 nm in radial step sizes of 0.003 cm. SEDNTERP⁴⁴ was used to estimate the partial specific volume of all UGDH constructs, and the buffer density (1.00726 g/ml) and viscosity (0.01018 P). SEDFIT⁴⁵ was used to model and fit all data. Data were modelled as a continuous sedimentation coefficient ($c(s)$) distribution. The baseline, meniscus, frictional coefficient, and systematic time-invariant, and radial invariant noise were fit⁴⁶. HYDROPRO⁴⁷ was used to predict s values based on crystal structures. The expected drag from the ID-tail was estimated by calculating the expected s values from crystal structures with and without modelled, energy minimized ID-tails. The data fits for all experiments can be found in Extended Data Fig. 3.

Evolutionary rate analysis. Seventy-nine UGDH sequences from vertebrates were used for analysis after removing redundancy at the organism level (only one UGDH sequence used per organism). The protein sequences were aligned using MUSCLE⁴⁸, and rates of evolution at each alignment position was calculated under the JTT model⁴⁹ using MEGA7 (log-likelihood method)⁵⁰. The rates were normalized such that the average rate of evolution was 1.0 across the entire protein. Residue positions evolving faster than average show a rate greater than 1.0. In Extended Data Fig. 4, only the rates at alignment positions where the human UGDH did not have an indel were used.

Monte Carlo sampling. The free-energy cost of tethering an unstructured, non-interacting peptide to an impenetrable surface depends on the ratio of all constrained and unconstrained states:

$$\Delta G_{\text{constrained}} = -RT \ln \left(\frac{\Omega_2}{\Omega_1} \right) \quad (1)$$

Where R is the gas constant, T is temperature, Ω_1 is the number of all possible states of an unconstrained, self-avoiding peptide and Ω_2 is the number of the Ω_1 states that do not conflict with the constraint imposed by the protein surface. To simplify, we used polyserine peptides, ignored side-chain entropy and used a hard sphere potential along with 166 coarse-grained ϕ , ψ bins to calculate Ω_1 and Ω_2 . Each bin represents a $10 \times 10^\circ$ range of ϕ , ψ values of peptide conformations in the 'allowed' region of the original Ramachandran map (Extended Data Fig. 5a, b). This calculation is nontrivial for large polymers, and an exhaustive grid search of all conformations was only conducted for the 3- and 4-residue ID-tails (Extended Data Fig. 5c). We used the following Monte Carlo procedure to estimate the fraction of surface-constrained conformations (Ω_2/Ω_1) for each ID-tail. To determine

the self-avoiding Ω_1 mesostates, we randomly assigned one of the 166 ϕ , ψ bins to each ϕ , ψ torsion angle in the ID-tail and then looked for steric clashes within the conformer using the 'outer limit' for atomic clashes as described in the original Ramachandran map⁵¹. Next, each of Ω_1 mesostates was analysed for steric clashes with the surface or the adjacent ID-tail (Extended Data Fig. 5d–l). Prior to the simulation, hydrogens were added to the hexamer structure using the 'reduce' program⁵², and an adjacent ID-tail was modelled in an extended conformation and fixed during the simulation (Extended Data Fig. 5d–f). The simulation was stopped when a minimum of 124,000 self-avoiding conformers were analysed and the ratio of surface-constrained conformations (Ω_2/Ω_1) reached convergence (Extended Data Fig. 5c). The convergence threshold was defined as a change in the cumulative ratio of less than 10^{-5} within a window of 5,000 trials. All runs reached convergence except for the 10-mer simulations, which only converged to 2 decimal places (Extended Data Fig. 5c–l). We estimated the accuracy in our Monte Carlo simulations by comparing the results to the full grid search of the 3- and 4-residue ID-tails (Extended Data Fig. 5c).

Thermodynamic shift assay. Solutions of UGDH (FL-dimer or Δ ID-dimer) at 0.1 mg/ml were prepared with 5 \times SYPRO Orange ThermoFluor (Thermo Fisher) in the standard reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA). Samples were then briefly spun and allowed to equilibrate for 20 min. The thermal denaturation experiments were conducted in replicates ($n \geq 3$) and data were acquired using a Bio-Rad MiniOpticon Real-Time qPCR machine. A fluorescence excitation spectrum wavelength between 470–505 nm and an emission spectrum between 540–570 nm were used. The fluorescence emission for each solution was recorded every 30 s as the temperature was increased from 25 to 80 °C (ramp speed of 0.5 °C/s). Baselines were subtracted from the raw data using the buffer control experiments. The baseline, plateau and slope of the denaturation curve were fit to equation (9) to obtain the apparent T_m (melting temperature) values⁵³.

$$Y = \text{baseline} + \frac{\text{plateau} - \text{baseline}}{1 + 10^{\frac{T_m - X}{\text{Slope}}}} \quad (9)$$

where Y represents the fluorescence signal at temperature X .

Hydrogen–deuterium exchange–mass spectrometry. Studies have shown that hydrogen–deuterium exchange (HDX) is an appropriate probe for protein dynamics and can illuminate differences between wild-type and mutant proteins^{54,55}. HDX is a powerful tool for footprinting the solvent-accessible regions of a protein⁵⁶, and was used in this study to compare structural and dynamic changes between the dimerized versions of UGDH (UGDH(FL-dimer) and UGDH(Δ ID-dimer)).

Proteins were expressed and purified in the Wood laboratory as previously described^{22–27}. Proteins were then flash-frozen and shipped overnight on dry ice to the Gross laboratory at Washington University in St. Louis for hydrogen–deuterium exchange–mass spectrometry (HDX–MS) analysis. Protein solutions (2 μ l) were continuously labelled at 25 °C by adding 20 μ l of 10 mM HEPES buffer containing 99.9% deuterium oxide ($pD = 7.4$). Samples were quenched by adding 33 μ l of 8 M guanidine hydrochloride and 100 mM TCEP (final pH = 3.0) at 30 s, 1 min, 2 min, 15 min, 1 h, and 2 h time points^{57,58}. One minute after quenching, samples were flash-frozen in liquid nitrogen and stored for less than 36 h at -80 °C. Control samples contained 10 mM HEPES in water rather than deuterium oxide. Each sample was thawed seconds before liquid chromatography followed by mass spectrometry (LC–MS). On-line protein digestion was performed with a custom-packed pepsin column (2 mm \times 20 mm) at a flow rate of 200 μ l/min in 0.1% trifluoroacetic acid. For desalting, a Zorbax Eclipse XDB-C-8 trap column (2.1 \times 15 mm, 3.5 μ m) was used to trap peptic peptides for 3 min. Following this, peptides were separated using a Hypersil Gold C-18 analytical column (2.1 \times 50 mm, 2.5 μ m), 4–80% gradient of acetonitrile with 0.1% formic acid (B), and a 100 μ l/min flow rate. Peptides were detected using a LTQ XL Orbitrap mass spectrometer (Thermo Fisher Scientific), with a mass resolving power of 50,000, m/z 400. Additional parameters were spray voltage of 5 kV, capillary temperature of 275 °C, capillary voltage of 49 V, and a tube lens of 163 V. All experiments were conducted in quadruplicate.

As a prelude to HDX, protein mapping was performed by identifying pepsin-digested peptides. Product-ion mass spectra were collected in the data-dependent mode, picking the six most abundant ions from selected MS/MS. Peptides were identified using Mascot (Matrix Science). Following HDX, mass spectra were analysed with HDX Examiner (Sierra Analytics). The per cent deuterium uptake was plotted against time for UGDH(FL-dimer) and UGDH(Δ ID-dimer). To magnify slight, yet significant changes in uptake, the cumulative differences in HDX for UGDH(FL-dimer) versus UGDH(Δ ID-dimer) were calculated. These values were plotted alongside 3 times the error propagation for all measurements of both variants for each peptide, after the data and error were normalized—divided by the number of time points considered for each data point (Extended Data Fig. 6). The propagation error for each peptide is equal to the square root of the sum

of all squared standard deviation values for collective time-dependent measurements of UGDH(FL-dimer) and UGDH(Δ ID-dimer). The cumulative per cent deuterium uptake was compared to 3 times the propagation error. Differences that were greater than 3 times the propagation error were noted as regions of change affected by the presence of the ID-tail. We chose to normalize the data to be more inclusive of peptides with low intensity that are found at most time points. In a similar manner, we have excluded those peptides that have avoided detection for more than two time points.

Stopped-flow analysis of UGDH hysteresis. The allosteric activation (E^* to E) of UGDH can be observed as a lag (hysteresis) in progress curves^{22,27} (see Extended Data Fig. 7f for examples). The allosteric activation rates for UGDH(FL) ($n \geq 6$) and UGDH(Δ ID) ($n \geq 6$) were monitored at 25 °C using an Applied Photophysics SX20 stopped-flow spectrophotometer. Enzyme solutions contained 500 nM UGDH(FL) or UGDH(Δ ID) in the standard reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA). This solution was rapidly mixed with an equal volume of standard reaction buffer that contained both substrate and cofactor. The mixed solution contained 250 nM UGDH(FL) or UGDH(Δ ID), with saturating amounts of both substrate and cofactor. The progress of the reaction was monitored by NADH production, with the absorbance reading at 340 nm being acquired every 10–15 ms. Progress curves were fit to equation (4) to determine the length of the lag in enzyme activation (E^* to E). The mean and standard deviation of the hysteretic lags were derived from 6 or more progress curves.

Transient-state kinetics of UDP-Xyl binding. Stopped-flow fluorescence studies were conducted at 25 °C using an Applied Photophysics SX20 stopped-flow spectrophotometer with a dead time of ~ 1.2 ms. Syringes were loaded with 500 nM UGDH(FL) or UGDH(Δ ID) and variable concentrations of UDP-Xyl, and then rapidly mixed. The change in intrinsic tryptophan fluorescence was continuously monitored using an excitation wavelength of 290 nm and an emission filter with a cut-off below 320 nm (Extended Data Fig. 7). Fluorescence decay curves were averaged from experimental replicates ($n \geq 4$) for each concentration in the series. Raw data was corrected for the inner-filter effect using the molar absorptivity at both the excitation and emission of UDP-Xyl⁵⁹. Data were globally fit using computer simulation with KinTek Global Kinetic Explorer program^{60,61} (KinTek). Multiple input models based on the known structural states were tested, and the best fit model was determined using confidence contour analysis⁶². Microscopic rate constants and errors are reported in Extended Data Fig. 7e. Fit data and confidence contours can be found in Extended Data Fig. 7a–d.

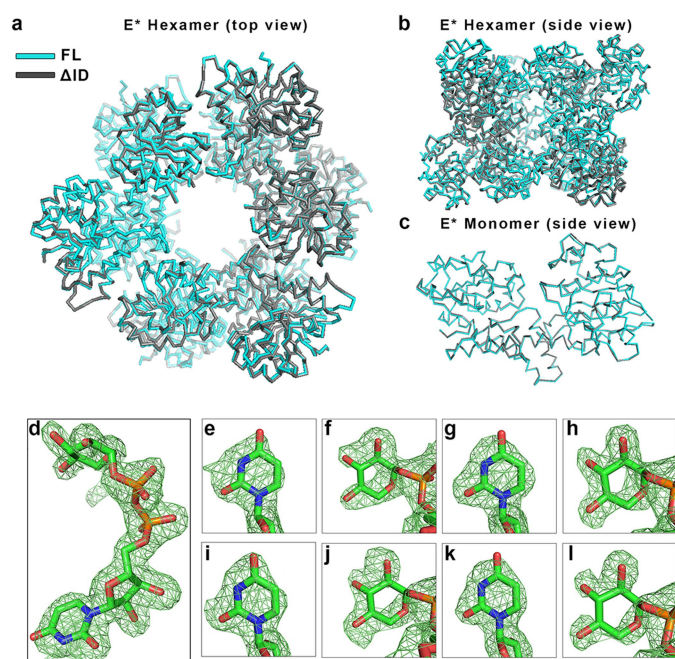
Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

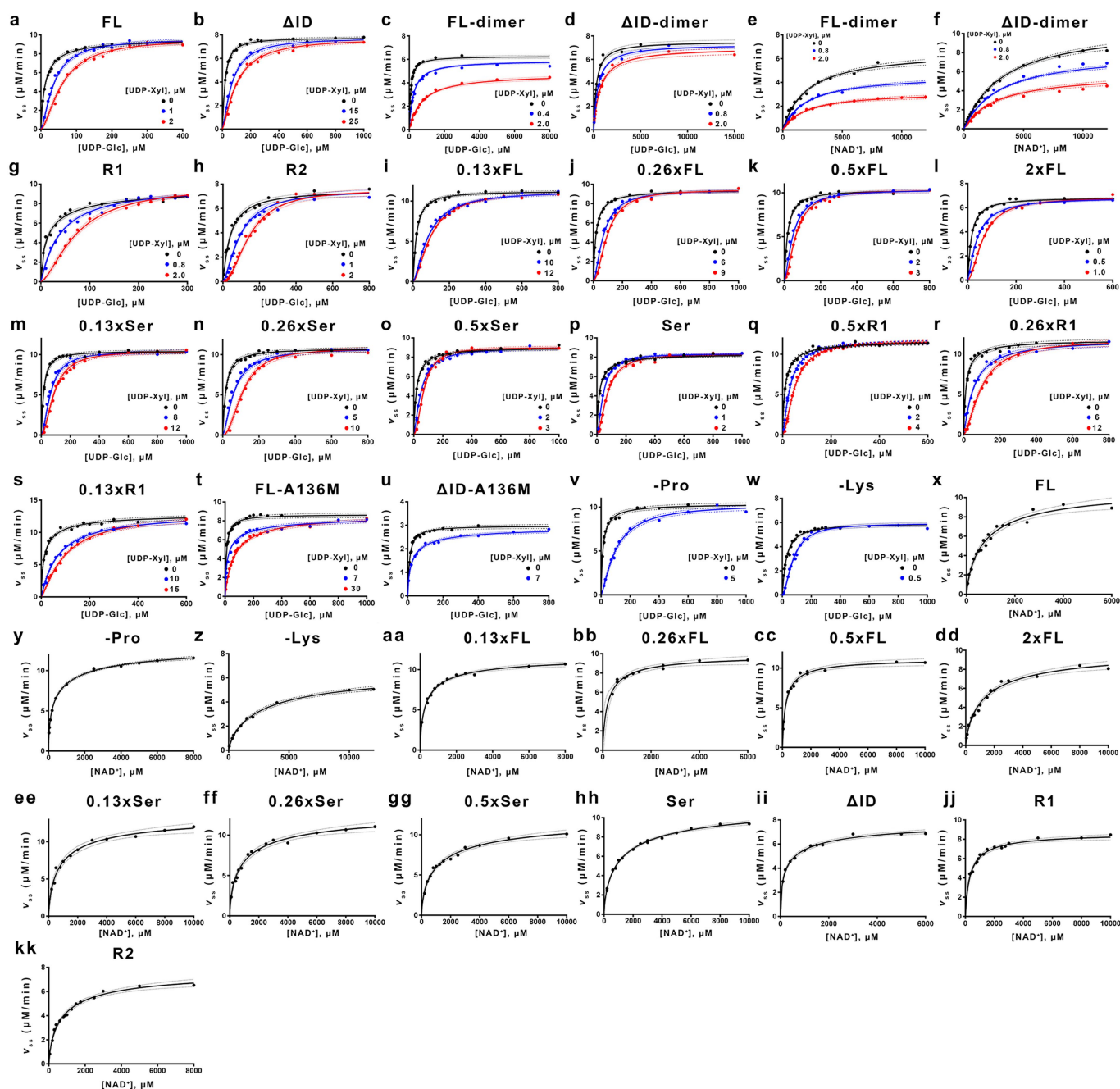
The structure factors and coordinates described in this manuscript have been deposited at the Protein Data Bank under accession codes 5W4X and 5VR8. All data generated or analysed in this study can be found in the Extended Data and the provided Source Data files.

- Wilkins, M. R. et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999).
- Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Brunger, A. T. Free R value: cross-validation in crystallography. *Methods Enzymol.* **277**, 366–396 (1997).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Urzhumtsev, A., Afonine, P. V. & Adams, P. D. TLS from fundamentals to practice. *Crystallogr. Rev.* **19**, 230–270 (2013).
- Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
- Diederichs, K. & Karplus, P. A. Improved R -factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269–275 (1997).
- Frieden, C. Kinetic aspects of regulation of metabolic processes. The hysteretic enzyme concept. *J. Biol. Chem.* **245**, 5788–5799 (1970).
- Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. *Analytical Ultracentrifugation*. (Royal Society of Chemistry, London, 1992).
- Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
- Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124 (2003).
- Ortega, A., Amoros, D. & Garcia de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101**, 892–898 (2011).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).

50. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
51. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
52. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999).
53. Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. Protein Sci.* **79**, 28.9.1–28.9.14 (2015).
54. Fang, J. et al. Conformational dynamics of the *Escherichia coli* DNA polymerase manager proteins UmuD and UmuD'. *J. Mol. Biol.* **398**, 40–53 (2010).
55. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158–170 (2006).
56. Johnson, B. et al. Dimerization controls Marburg virus VP24-dependent modulation of host antioxidative stress responses. *J. Mol. Biol.* **428**, 3483–3494 (2016).
57. Chen, E. et al. Broadly neutralizing epitopes in the *Plasmodium vivax* vaccine candidate Duffy Binding Protein. *Proc. Natl Acad. Sci. USA* **113**, 6277–6282 (2016).
58. Yan, Y., Grant, G. A. & Gross, M. L. Hydrogen–deuterium exchange mass spectrometry reveals unique conformational and chemical transformations occurring upon [4Fe-4S] cluster binding in the type 2 L-serine dehydratase from *Legionella pneumophila*. *Biochemistry* **54**, 5322–5328 (2015).
59. Palmier, M. O. & Van Doren, S. R. Rapid determination of enzyme kinetics from fluorescence: overcoming the inner filter effect. *Anal. Biochem.* **371**, 43–51 (2007).
60. Johnson, K. A. Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol.* **467**, 601–626 (2009).
61. Johnson, K. A., Simpson, Z. B. & Blom, T. Global Kinetic Explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387**, 20–29 (2009).
62. Johnson, K. A., Simpson, Z. B. & Blom, T. FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem.* **387**, 30–41 (2009).

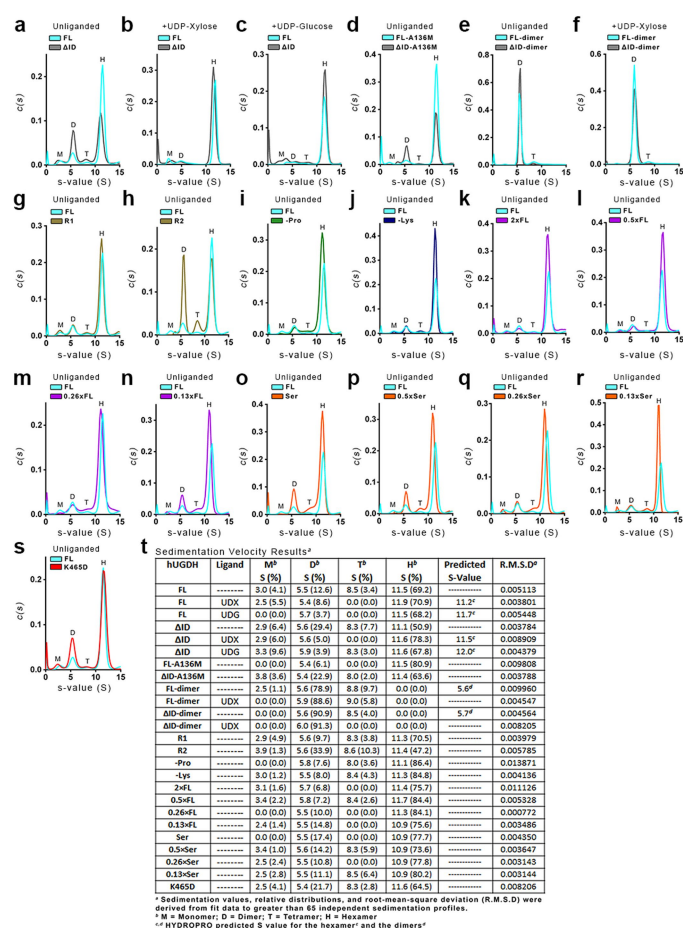


Extended Data Fig. 1 | The crystal structures of UGDH(FL) and UGDH(Δ ID) show no significant structural deviations, and structural evidence for UDP-Xyl binding in the NAD⁺ site. a–c, Structural overlay (root mean square deviation (r.m.s.d) = 0.385 Å), comparing the UGDH(FL) (cyan) and UGDH(Δ ID) (grey) E* hexamers (a, b) and monomers (c). PDB entries for UGDH(FL) and UGDH(Δ ID) are 4RJ7 and 5W4X, respectively (Extended Data Table 1). d, Crystal structure of native UGDH with UDP-Xyl bound in the active site. Difference density map ($F_0 - F_c$) of UDP-Xyl (chain B) calculated at 2.0 Å resolution and contoured at 3.5 σ . The map was calculated after omitting the UDP-Xyl and subjecting the model to simulated annealing. e–l, UDP-Xyl can also bind weakly to the NAD⁺-binding site of native UGDH. Difference electron density maps ($F_0 - F_c$) were calculated as in d. The uracil and xylose in the NAD⁺-binding sites were contoured at 3.5 and 3 σ for chain A (e and f, respectively), chain B (g and h, respectively), chain D (i and j, respectively) and chain E (k and l, respectively). Chains C and F do not contain UDP-Xyl in the NAD⁺-binding site. UDP-Xyl binding in the NAD⁺ site is the source of mixed inhibition observed in the UGDH(FL-dimer) and UGDH(Δ ID-dimer) constructs. (see Supplementary Information, Section 1). PDB entry: 5VR8 (this work, Extended Data Table 1).

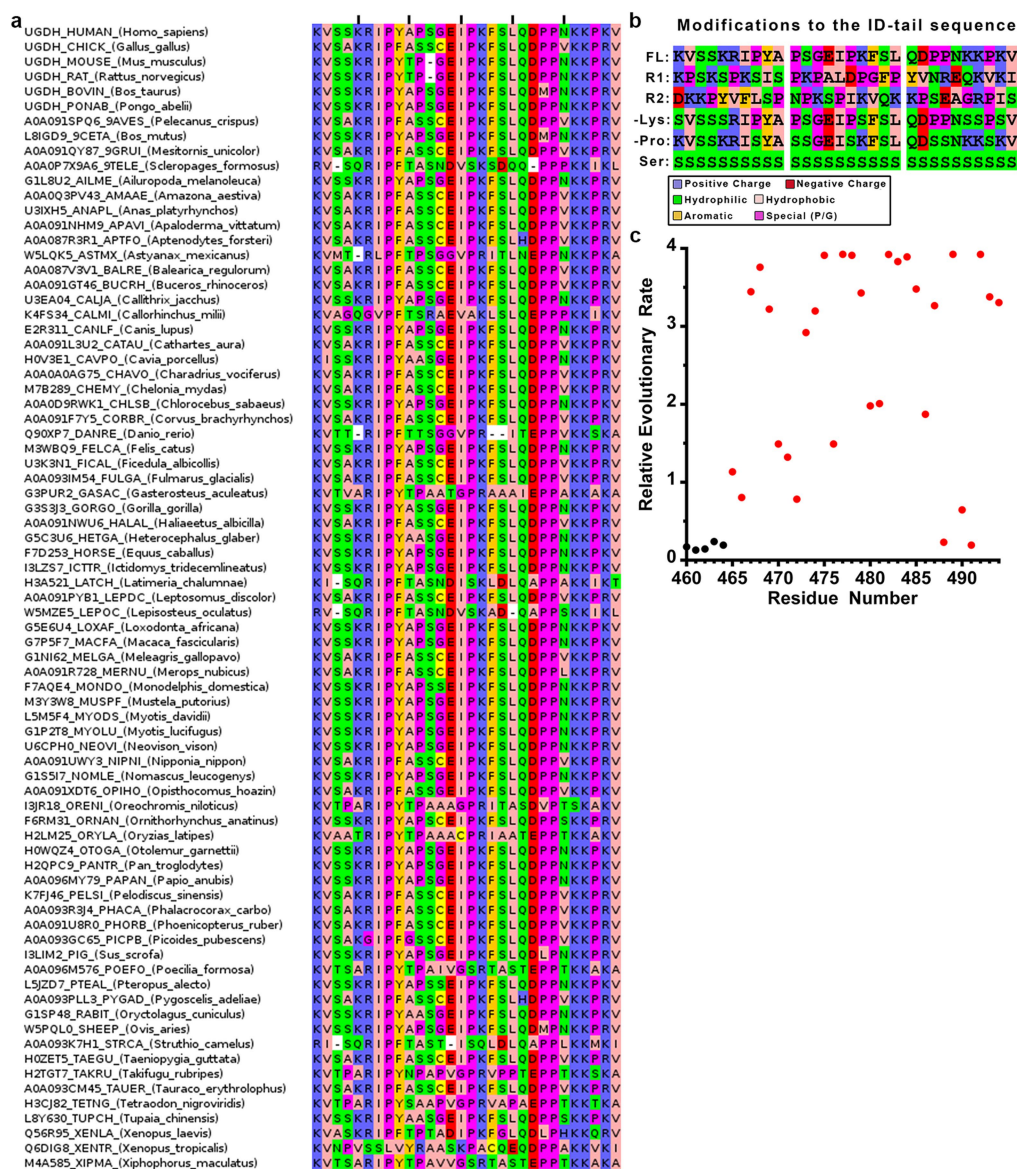


Extended Data Fig. 2 | Steady-state kinetic analysis of all UGDH constructs. a–w, Inhibition studies with the allosteric inhibitor UDP-Xyl. Data from two or three independent rate curves were globally fit to equation (7) (or equation (8) for dimers c–f) using nonlinear regression ($n \geq 26$ data points). See Extended Data Table 2 for the specific number

of data points and fit parameters. Dashed lines indicate 95% confidence intervals. **x-kk**, NAD⁺ substrate-saturation curves fit to equation (6) using nonlinear regression ($n \geq 10$ independent data points). See Extended Data Table 3 for the specific number of data points used in global fitting.



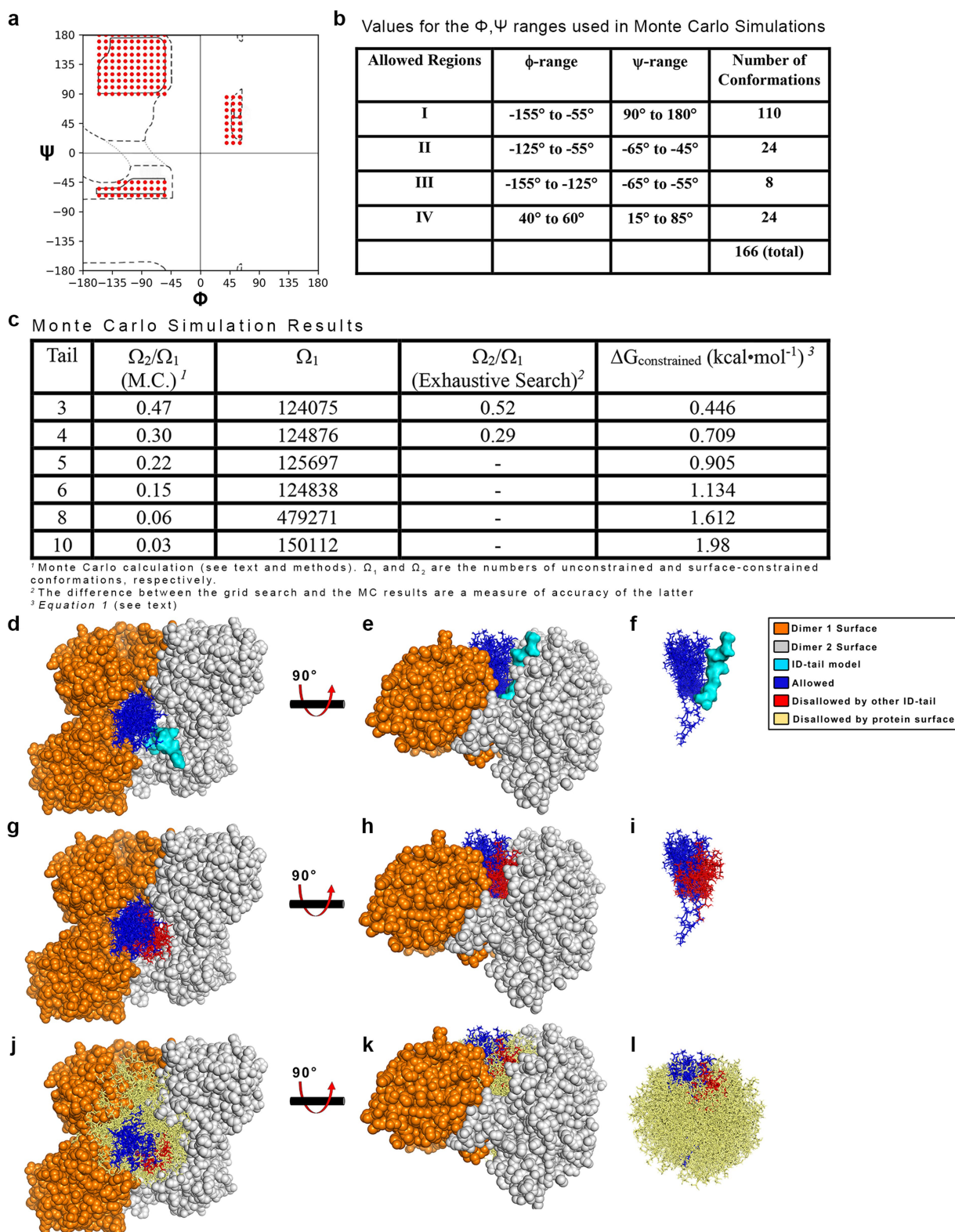
Extended Data Fig. 3 | Sedimentation velocity studies of the UGDH constructs. **a–s.** Plots of the $c(s)$ distributions with oligomeric species labelled as H (hexamer), T (tetramer), D (dimer) or M (monomer). The R2 mutant (**h**) shows no change in UDP-Xyl affinity (Fig. 2c and Extended Data Table 2), yet shows evidence of a less stable hexamer. Panel **s** was included to show that the hexamer in **h** is less stable partly owing to the K465D substitution in the UGDH(R2) construct. The K465D substitution introduces an unfavourable negative charge near E460 in the hexamer interface, which may reduce the stability. **t.** Relative distributions, s values (S) and r.m.s.d. values for all sedimentation velocity experiments.



Extended Data Fig. 4 | The ID-tail is conserved in vertebrates.

a, ClustalO sequence alignment of all vertebrate UGDH ID-tail regions (79 total). Residues are coloured by type, where blue is positive charge (K, R, H), red is negative charge (D, E), peach is hydrophobic (A, V, L, I, M), orange is aromatic (F, W, Y), green is hydrophilic (S, T, N, Q), yellow is cysteine (C), and magenta is special (P, G). **b**, The ID-tail was extensively randomized and modified. Sequences of UGDH (FL, R1, R2, -Lys, -Pro,

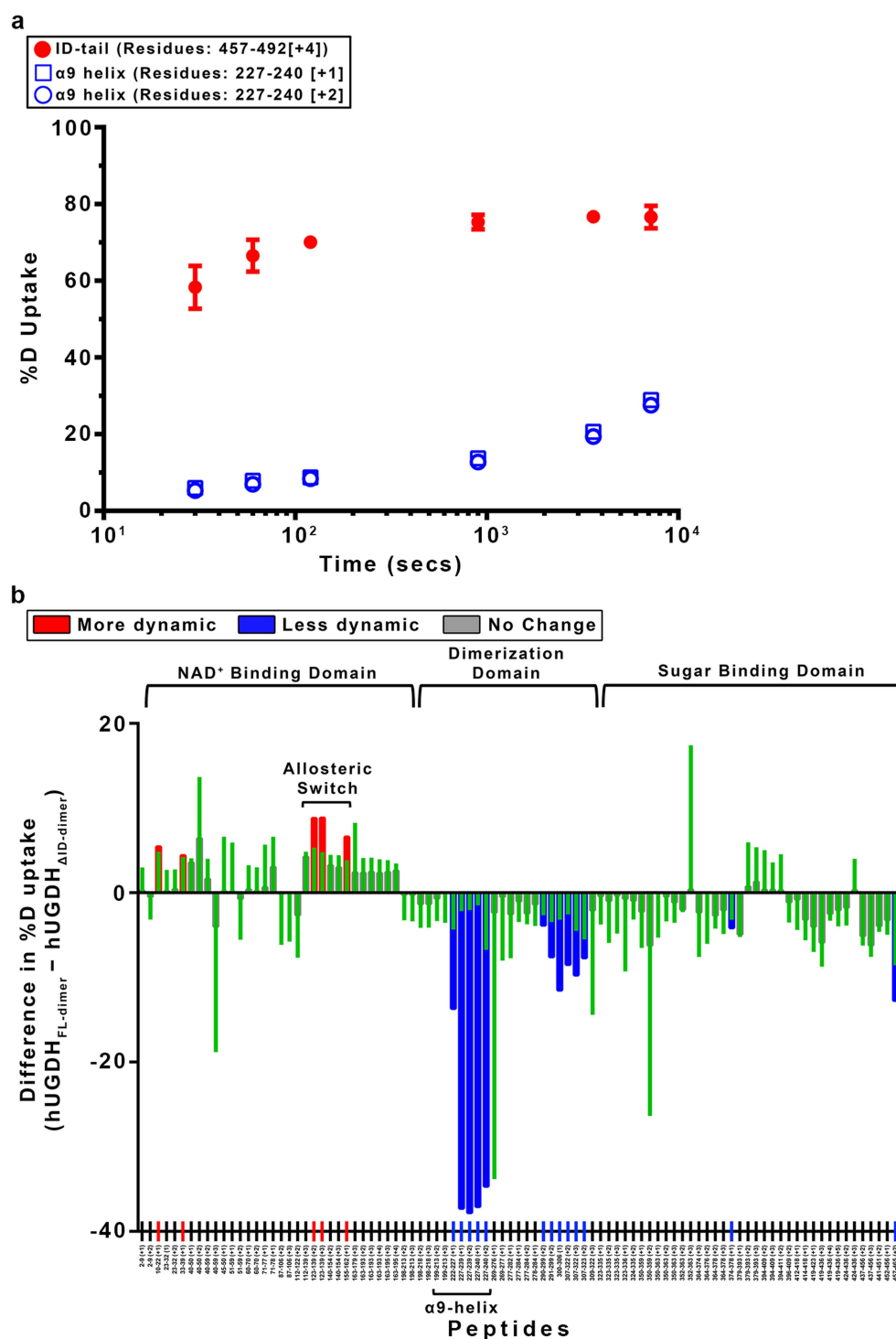
and Ser), aligned by position and coloured by residue type. **c**, Relative evolutionary rate of UGDH residues from the alignment of 79 vertebrate sequences. The ID-tail (red dots) begins at residue 465 and displays an approximately threefold higher rate of divergence than the folded portion of the protein (black dots). For clarity, only a small, representative segment of the folded protein is shown (residues 460–464). All rates were scaled such that the average rate is 1.0 across the entire dataset.



Extended Data Fig. 5 | Exhaustive Monte Carlo simulations

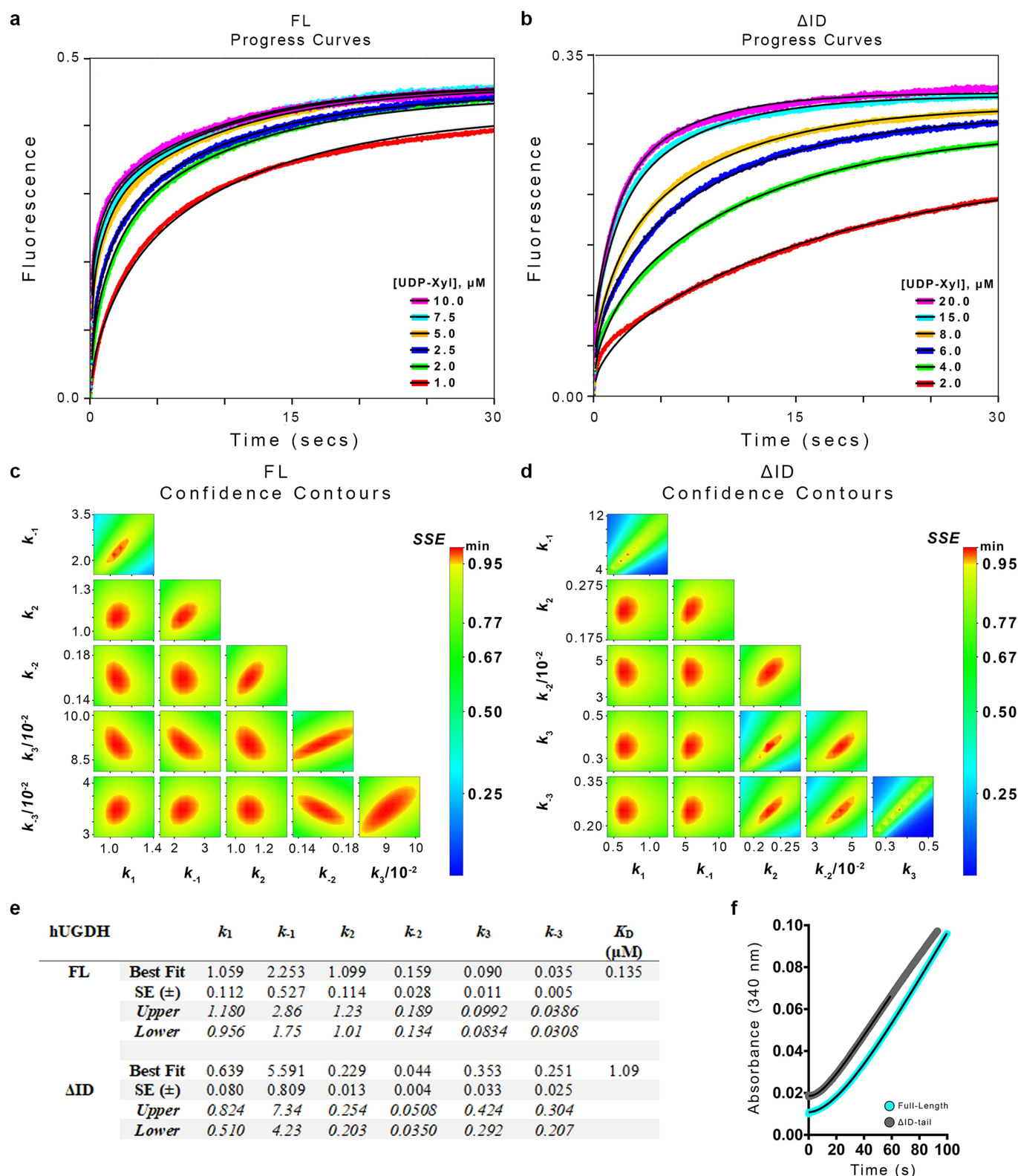
constraining the ID-tail. **a**, Dashed lines outline the traditional, generously allowed regions of the Ramachandran plot, whereas the red circles identify the conformations used in the Monte Carlo simulations. **b**, The ranges of ϕ and ψ angles depicted in **a**. The $10 \times 10^\circ$ bins are centred on the first and last numbers in the range. For example, in region I, the first ϕ, ψ bin ($-155^\circ, 90^\circ$) represents the ϕ range -155° to -145° and the ψ range 85° to 95° . **c**, Ratio of ID-tail conformations constrained (Ω_2) to the number of conformations when the ID-tail is unconstrained (Ω_1). The entropic costs of confining tails of each length were calculated

using equation (1). **d, e**, The results of the 10-residue ID-tail simulations, shown in a surface representing the hexamer-building interface (orange and grey dimers) with the adjacent ID-tail (cyan) that was fixed during simulations. Also depicted is a representative sampling of 20 allowed Ω_2 conformations (blue sticks) from the 4,503 identified in the Monte Carlo simulation. **f**, The same view as in **e**, but without the protein surface. **g–i**, Same as in **d–f**, but now including a sampling of 20 of the 3,002 Ω_1 conformations (red sticks) that clash with the fixed adjacent ID-tail (not depicted for clarity). **j–l**, Same as **g–i**, but including 750 of the 142,607 Ω_1 conformations (tan sticks) that clash with the protein surface.



Extended Data Fig. 6 | The ID-tail induces global changes in the structure and dynamics of UGDH. a, The per cent deuterium uptake of the ID-tail peptide region (residues 457–492; red closed circles) saturates rapidly, consistent with an unfolded peptide³¹. For comparison, two peptides corresponding to the well-ordered α 9 helix region (open blue squares and circles) saturate slowly. Data are mean \pm s.d. of independently replicated time points ($n = 4$). For some points, the standard deviation is less than the dimensions of the data symbol. **b,** The normalized cumulative changes in the hydrogen–deuterium exchange rates (UGDH(FL-dimer) –

UGDH(Δ ID-dimer)). Most of the kinetics measurements consisted of six independently replicated time points ($n = 4$), processed to give the mean exchange (red, blue or green bars). Approximately 5% of the data displayed low signal:noise or was missing, and in those cases the means were derived from four or more time points. Results were normalized by dividing by the number of measurements. The propagation error for each peptide is equal to the square root of the sum of all squared standard deviation values for the collective measurement of UGDH(FL-dimer) and UGDH(Δ ID-dimer).



Extended Data Fig. 7 | Transient-state analysis of UGDH(FL) and UGDH(Δ ID). **a, b**, Transient-state analysis of UDP-Xyl binding kinetics using intrinsic protein fluorescence. Six independent progress curves (coloured traces) at different inhibitor concentrations were globally fit (black line) to the allosteric inhibition model (see Fig. 4b) for UGDH(FL) and UGDH(Δ ID). Each progress curve was replicated ($n \geq 4$) with similar results, and the final kinetic model was refined against the averaged progress curves (see **e** for fit parameters). **c, d**, Confidence contour plots depicting how constrained each globally fit parameter is relative to the others, for all progress curves in **a** and **b** (parameters are listed in **e**). **e**, Table of the microscopic rate constants from global fitting of the

progress curves described in **a** and **b**. The best fit and s.e.m. were obtained from global nonlinear regression based on the numerical integration of rate equations for the described model (see main text and Methods). Upper and lower limits were obtained from the confidence contour analysis. $K_d = (K_1 K_2 K_3)^{-1}$, where $K_n = k_n / k_{-n}$. **f**, Enzyme hysteresis is observed as a lag in progress curves. Representative progress curves (of $n = 6$ independent measurements) for both UGDH(FL) (cyan) and UGDH(Δ ID) (grey) are fit to equation (4) (black line). Curves are displayed with the y axis offset for clarity. Final results for all replicate curves are displayed in Fig. 4a.

Extended Data Table 1 | Data collection and refinement statistics

Data collection		
Protein Data Bank Entry	5W4X E [*] hUGDH _{ΔID} C2	5VR8 E ^Δ hUGDH _{FL} P12;1
Space group		
Unit cell dimensions a,b,c (Å)	178.19, 114.07, 97.24 (116.9°)	89.08, 196.49, 111.26 (111.9°)
Completeness (%)	99.9 (91.1) ^a	93.2 (60.0) ^a
No. reflections	324,675	2,730,154
Redundancy	6.4 (6.1)	12.3 (10.3)
<i>I</i> / $\sigma(I)$	21.9 (1.5)	14.9 (2.5)
CC _{1/2} ^b	99.9 (64.9)	99.7 (79.3)
R _{meas} (%) ^c	6.5 (122.5)	13.2 (89.3)
Refinement		
Resolution (Å)	2.65	2.00
R _{work} / R _{free}	0.19 / 0.23	0.16 / 0.19
No. atoms: Protein / Ligand / Water	10887 / 33 / 36	21584 / 394 / 1097
B-factors (Å ²): Protein / Ligand / Water	89.9 / 97.4 / 64.3	33.2 / 27.1 / 32.3
Stereochemical Ideality		
Bond lengths (Å ²)	0.004	0.008
Bond angles (°)	0.75	0.91
φ,ψ Preferred (%) ^d	98.98	97.8
φ,ψ Additionally allowed (%)	1.02	2.2
φ,ψ Disallowed region (%)	0.0	0.0

^aValues in parenthesis are for the highest-resolution shell (2.71–2.64 and 2.0221–1.9994 for 5W4X and 5VR8, respectively).

^bCC_{1/2} is the percentage of correlation between intensities from random half-data sets⁴¹.

^cR_{meas} is the redundancy-independent merging R factor⁴².

Extended Data Table 2 | Kinetic parameters of all UGDH constructs^a

hUGDH	K_M (UDP-Glc, μM)	k_{cat}^b (s^{-1})	K_i^{UDX} (UDP-Xyl, μM)	α_{UDG}^c	$\Delta \Delta G^e$ ($\text{kcal}\cdot\text{mol}^{-1}$)	# of Data Points ^h
ΔID	17.8 ± 0.9	0.7 ± 0.01	5.44 ± 0.55	-----	0.00	42
FL	12.7 ± 0.6	0.8 ± 0.01	0.52 ± 0.04	-----	-1.39	38
R1	12.9 ± 1.0	0.8 ± 0.01	0.60 ± 0.06	-----	-1.31	59
0.13×R1	12.8 ± 1.2	1.0 ± 0.01	2.59 ± 0.24	-----	-0.44	40
0.26×R1	12.4 ± 1.0	1.0 ± 0.01	1.81 ± 0.18	-----	-0.65	42
0.5×R1	11.1 ± 0.8	1.0 ± 0.01	1.09 ± 0.08	-----	-0.95	47
R2	43.7 ± 3.6	0.7 ± 0.01	0.78 ± 0.07	-----	-1.15	50
-Lys	30.1 ± 1.9	0.5 ± 0.01	0.29 ± 0.03	-----	-1.73	39
-Pro	13.1 ± 0.9	0.9 ± 0.01	0.72 ± 0.07	-----	-1.20	26
0.13×FL	18.8 ± 0.9	1.0 ± 0.01	2.76 ± 0.15	-----	-0.40	46
0.26×FL	18.3 ± 0.7	0.8 ± 0.01	1.99 ± 0.12	-----	-0.60	42
0.5×FL	18.8 ± 0.9	0.9 ± 0.01	1.12 ± 0.08	-----	-0.94	50
2×FL	15.2 ± 0.7	0.6 ± 0.01	0.30 ± 0.02	-----	-1.72	43
0.13×Ser	16.9 ± 1.0	0.9 ± 0.01	2.67 ± 0.24	-----	-0.42	49
0.26×Ser	18.4 ± 1.0	0.9 ± 0.01	1.76 ± 0.18	-----	-0.67	43
0.5×Ser	17.4 ± 1.3	0.8 ± 0.01	1.09 ± 0.10	-----	-0.95	49
Ser	17.8 ± 1.0	0.7 ± 0.01	0.60 ± 0.05	-----	-1.31	53
ΔID-dimer	286 ± 27	0.1 ± 0.01	1.23 ± 0.15^d	22 ± 12	0.00^f	36
FL-dimer	83.2 ± 2.2	0.1 ± 0.01	0.17 ± 0.01^d	36 ± 5	-1.17^f	50
ΔID-A136M	9.9 ± 0.6	0.3 ± 0.01	4.20 ± 0.51	-----	0.00^g	30
FL-A136M	8.5 ± 0.6	0.7 ± 0.01	4.41 ± 0.37	-----	0.03^g	55

^aKinetic parameters and associated s.e.m. for all constructs were derived from global analyses of data in Extended Data Fig. 2.^bOne catalytic turnover of UDP-GlcA produces two molecules of NADH per cycle.^c α describes the mode of mixed inhibition (equation (8)). An $\alpha > 1$ in the UDP-Glc saturation curves shows that UDP-Xyl binds preferentially to the allosteric binding site, and secondarily to the coenzyme-binding site.^dCompetitive K_i from the fit to the mixed inhibition (equation (8)).^eChange in UDP-Xyl binding free energy (kcal mol^{-1}) of UGDH constructs relative to UGDH(ΔID): $\left(\Delta \Delta G = RT \ln \frac{K_i^{\text{Construct}}}{K_i^{\Delta\text{ID}}} \right)$.^fChange in UDP-Xyl binding free energy relative to the UGDH(ΔID -dimer).^gChange in UDP-Xyl binding free energy relative to the UGDH(ΔID -A136M).^hThe number of independent data points used in global analysis (see Methods).

Extended Data Table 3 | NAD⁺ kinetic parameters for UGDH^a

hUGDH	K_M (NAD ⁺ , mM)	$K_{0.5}^b$ (NAD ⁺ , mM)	Hill (h)	k_{cat}^c (s ⁻¹)	UDX (K_i , μ M)	α_{NAD}^d	# of Data Points ^e
FL	-----	0.8 ± 0.20	0.8 ± 0.1	0.9 ± 0.08	-----	-----	18
Δ ID	-----	0.3 ± 0.06	0.6 ± 0.1	0.7 ± 0.03	-----	-----	12
FL-dimer	2.0 ± 0.26	-----	-----	0.1 ± 0.01	2.1 ± 0.4	0.9 ± 0.2	37
Δ ID-dimer	3.2 ± 0.10	-----	-----	0.2 ± 0.01	3.6 ± 0.8	0.6 ± 0.2	47
R1	-----	0.4 ± 0.03	0.9 ± 0.1	0.7 ± 0.01	-----	-----	17
R2	-----	0.8 ± 0.14	0.7 ± 0.1	0.7 ± 0.01	-----	-----	15
-Lys	-----	2.9 ± 0.61	0.8 ± 0.1	0.6 ± 0.04	-----	-----	10
-Pro	-----	0.5 ± 0.06	0.6 ± 0.1	1.2 ± 0.03	-----	-----	12
0.13×FL	-----	0.4 ± 0.03	0.7 ± 0.1	1.0 ± 0.03	-----	-----	13
0.26×FL	-----	0.2 ± 0.03	0.8 ± 0.2	0.8 ± 0.05	-----	-----	11
0.5×FL	-----	0.3 ± 0.03	0.9 ± 0.1	0.9 ± 0.02	-----	-----	12
2×FL	-----	1.4 ± 0.31	0.8 ± 0.1	0.9 ± 0.01	-----	-----	18
0.13×Ser	-----	0.9 ± 0.24	0.7 ± 0.1	1.2 ± 0.10	-----	-----	12
0.26×Ser	-----	1.0 ± 0.27	0.7 ± 0.1	1.1 ± 0.09	-----	-----	15
0.5×Ser	-----	1.2 ± 0.34	0.7 ± 0.1	1.1 ± 0.09	-----	-----	13
Ser	-----	1.3 ± 0.19	0.7 ± 0.1	1.0 ± 0.04	-----	-----	15

^aKinetic parameters and associated s.e.m. for all constructs were derived from global analyses of data in Extended Data Fig. 2.

^bHexameric UGDH displays negative cooperativity with NAD⁺ binding, which indicates a mix of high-affinity and low-affinity sites²²⁻²⁷. In previous work, we showed that the native UGDH(FL) $K_{0.5}$ of 0.8 mM NAD⁺ corresponds to a mix of high-affinity and low-affinity sites with K_M of 88 μ M and 1.8 mM, respectively²³. This is consistent with the published K_d of 30 μ M for the coenzyme²³.

^cOne catalytic turnover of UDP-GlcA produces two molecules of NADH per cycle.

^d α describes the mode of mixed inhibition (equation (8)). An $\alpha < 1$ in the NAD⁺ saturation curves show that UDP-Xyl binds preferentially to the allosteric binding site, and secondarily to the coenzyme-binding site.

^eThe number of independent data points used in nonlinear regression (see Methods).

CAREERS

SHARE Tell us your career story at
naturecareerseditor@nature.com

E-NEWSLETTER Sign up for our careers digest
at go.nature.com/careersnewsletter

SOCIAL Follow us on Twitter at
twitter.com/naturejobs

NATHAN CHAPPELL



Members of Te Pūnaha Matatini Whānau, a student-led group in New Zealand, relish opportunities to build skills such as event planning and networking.

GRADUATE-STUDENT INITIATIVES

How advocacy gives back

Purpose-driven student leaders learn management skills and gain a support community.

BY KENDALL POWELL

In 2016, structural biologist Christina Roman co-founded a graduate-student drive to help recruit and retain a more diverse group of PhD candidates at her university. Since then, the team has boosted the numbers of successful PhD applicants from under-represented groups. She loved seeing how her team's work was directly affecting the lives of others.

Roman and two student colleagues at the University of Chicago in Illinois had set up the drive, known as the Graduate Recruitment Initiative Team (GRIT), to address a disconnect between the university's approach to diversity and the needs of graduate students from minority groups.

But for Roman, a PhD student in biochemistry and molecular biophysics, working with GRIT yielded another powerful benefit — organizational skills that she could apply

directly to her thesis work. To organize GRIT's efforts, she produced spreadsheets that listed completed tasks, approaches that had or hadn't worked and goals yet to be completed, broken down into 'next actions'.

She realized that she could use the same rigorous planning, note-taking and record-keeping habits in the lab, including printing out all her data and stapling them into a lab notebook so that she could have everything in front of her and in one place. "I organized my working space and my brain to fit how I was looking at recruitment and retention of diverse students," says Roman. "Now, I can see the fastest solution to try in my research."

She says that her advocacy work has also given her more confidence in pursuing her degree. "For me, as a person of colour, the PhD sometimes seems impossible. But my work in GRIT makes me feel like I can do it."

Although some PhD programmes offer training and workshops in organizational

skills, volunteer activism and initiatives give students hands-on experience with team and project management and negotiation, all of which transfer well to a science career.

Volunteering also creates opportunities for students to meet people outside their programme, offering an escape from the pressures of the lab. "The GRIT students," says Vicky Prince, dean of graduate affairs for the Biological Sciences Division at the University of Chicago, "have emerged as leaders".

MINDFUL CONVERSATIONS

Today, GRIT has 60 members and links applicants from all under-represented groups (URGs) with current biological- and physical-sciences students, who act as mentors during the application and recruitment process and connect new students with faculty mentors. "We saw the power in getting faculty members who aren't minorities to hear the stories and experiences of minority students in ►

► science,” says biochemist Cody Hernandez, another GRIT co-founder.

GRIT’s leaders say that they have perfected the crucial skills of negotiating and listening. These qualities are especially important during the sometimes-tense conversations they have with faculty members about race and self-identity, and with colleagues who hold strong and conflicting opinions. The students have worked, for example, to refute the false idea that few qualified URG students apply to their university, and have shared with majority-group faculty members how negative stereotypes affect their career advancement.

“You are going to hear things you don’t like from people you respect,” says biochemist and third GRIT co-founder Mathew Perez-Neut. “You can’t be inflammatory or aggressive. You have to empathize, be tolerant and be patient in these situations.”

He thinks that these skills will translate well to managing teams of researchers who won’t always agree on scientific approaches or interpretations of data. “GRIT continually gives back to me,” Perez-Neut says, citing the strong relationships he has forged with faculty members — who will write his recommendation letters — and student peers.

But he also stresses the importance of not taking on too much as a volunteer, and of sharing the burden with co-pilots who can take over when research gets busy or life overwhelms. It’s equally important, he adds, to have faculty-member allies who can offer advice, help students to navigate administrative bureaucracy and advocate on behalf of a student group.

People often expect more from student leaders than from students who are less involved, and so ask them to participate in other activities. Those who launch or join such volunteer initiatives must learn to say ‘no’ when an activity doesn’t line up closely with the original goal, the GRIT leaders advise. “It’s important to have a laser-beam focus on what you want to accomplish,” Perez-Neut says.

During his master’s degree in ecology and evolution at the École Normale Supérieure (ENS) in Paris, Lucas Paoli focused on how students could address climate change. As part of a French national student sustainability group, Paoli had the opportunity to participate in the United Nations Framework Convention on Climate Change Conference of the Parties (COP) meetings. At these events, countries negotiate and implement climate-change policies.

But he noticed that few French universities were participating in the meetings, even though they are encouraged to join in as research non-governmental organizations (RNGOs). So, Paoli proposed that the ENS apply for COP accreditation to participate in the UN meetings as an RNGO.

Working with Christian Lorenzi, director of scientific studies for the ENS, Paoli secured accreditation last July for delegates from his university to attend the COP-24 meeting in Katowice, Poland, which will take place next month,

and other COP meetings beyond. The two also designed a course to prepare student attendees for the experience. “These COP meetings are where the international climate-change regime is created,” says Paoli. “Sending students seemed the best way to train them for such fields.”

Lorenzi, a cognitive neuroscientist, says that the programme’s early success is due to Paoli’s passion, which convinced him that ENS students should participate in the COP events.

“It’s important to have a laser-beam focus on what you want to accomplish.”

He says that Paoli was highly organized and efficient: “He had the capacity to engage me with his confidence that it was possible to do this big thing.”

For his part, Paoli says that his COP experiences helped to prepare him for interviews and to secure his current position as a PhD student in ocean microbiology at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

His efforts have also provided an outlet for his creative energy when he needs to do something other than lab work. “I can’t write bioinformatics code for the entire day,” he adds.

PURPOSEFULLY CONNECTED

A student-led group through New Zealand’s Te Pūnaha Matatini (TPM) Centre of Research Excellence, hosted by the University of Auckland, helps to foster its members’ event-planning and time-management skills. Called TPM Whānau, the group grew out of a desire to connect the centre’s 70 or so early-career researchers who are spread across the country, enabling transdisciplinary research on complex systems and networks.

‘Whānau’ translates roughly to ‘extended family’ in the Indigenous language of New Zealand, te reo Māori. The group’s events help to create a sense of community among members because they are geographically distant from each other, explains Reno Nims, an anthropology PhD candidate at the University of Auckland and current chair of TPM Whānau.

The TPM Whānau committee meets twice monthly on Skype or Zoom videoconference calls. In 2017, the group held a retreat that involved all facets of event planning — booking a meditation centre in the hills an hour outside of Auckland, organizing a hike into the bush and arranging speakers on topics such as big data, privacy and data sovereignty among Indigenous groups.

For student leaders who are juggling research, classes and extracurricular activities, time management is essential. Nims navigates his lab work and TPM Whānau responsibilities using the ‘Pomodoro’ method, setting a timer for 25 minutes of uninterrupted work on a particular task, followed by 5 minutes of break. He uses the project-management app Trello to break down a big event into individual tasks, and LeechBlock to selectively block web-surfing.

He reserves Mondays and Fridays for e-mails, literature reading and TPM Whānau projects. On Tuesdays, Wednesdays and Thursdays, he sifts through and identifies fish bones from historic Māori archaeological sites for his thesis research.

Nims’s days are busy, but he says that his motivation to be a student leader arises partly from a wish to build a research community for himself. “I’m creating an environment where I can get something out of it — that keeps up my energy to keep doing it,” he says.

At Washington University in St Louis, Missouri, meanwhile, student leaders in the BioEntrepreneurship Core (BEC) help other students to gain entrepreneurship knowledge and relevant skills.

The BEC hosts seminars or meetings with chief executives and founders of start-up biotechnology or biopharmaceutical companies. “These are a bit like informational interviews (where potential jobseekers ask for informal career advice), but with a dozen of us — everyone benefits from being able to ask a chief executive questions,” says Zuzana Kocsisova, co-president of the BEC and a PhD candidate in developmental biology.

The BEC also organizes tours of local companies and holds an annual business-planning competition.

Initially, Kocsisova joined as the BEC marketer, enabling her to learn basic graphic-design skills in Adobe Illustrator, how to build a website, handle e-mail-marketing lists and use LinkedIn to invite speakers. “These are things I would not have learned as a grad student, and I learned them very fast,” she says.

She also learnt how to delegate effectively by asking a specific person to do a task by a clear deadline, scrapping her earlier practice of throwing out a group e-mail that asked: “Can anyone do this?”

Kocsisova also boosted her networking skills by building connections with the St Louis entrepreneur community. Last year, for example, she attended a fundraiser and talked with the city’s major venture-capital investors.

Kocsisova says that she views working with the BEC, or on other student-advocacy activities, as an integral part of postgraduate training. The activities aren’t taking away from anyone’s studies, she says: “Becoming a leader is what you are supposed to be doing in graduate school.” ■

Kendall Powell is a freelance science writer based in Lafayette, Colorado.

CORRECTION

The Spotlight article ‘An alternative Japan experience’ (*Nature* **562**, S53–S55; 2018) incorrectly stated that Nak Young Chong joined JAIST as a visiting professor. In fact, he was jointly appointed as associate professor at both JAIST and AIST.

HOW IT FEELS TO BE SWALLOWED BY A BLACK HOLE

Into the unknown.

BY GRETCHEN TESSMER

How does it feel to be swallowed by a black hole? The short answer is: not great.

I suppose there's less stretching than I expected — no rapid, violent pulling apart of life and limb. The ship remains in one piece. So do I. So does Emma, who really deserves to be ripped apart, after skimming us so close to the edge of this black hole in the first place. And Julius, who deserves it even more, because he's always been the one to rein in Emma's recklessness before, and we've come to depend on it. This time around, he failed completely.

But these two are endlessly disappointing. After years of bickering, baiting each other, and indulging in some classic pilot-navigator sexual tension that had the rest of us rolling our eyes and covering our ears, they finally started hooking up on the last trip. Whatever. I don't begrudge the whole 'falling in love' nonsense... unless it leads to me cresting the edge of a black hole and slipping down the other side with the sticky, slow sensation of molasses flowing downhill.

What a god-awful mess.

They say time slows when you enter a black hole. Three hundred thousand years can pass by in three minutes. Maybe. But it certainly doesn't feel like eternity. Unless eternity only needs three minutes to wrap itself up? The whole notion seems terribly anticlimactic.

We stand at the rear window of our ship, all six of us — Emma, Julius, Myreen and the shadow-frequency, Lex Bethel and me — all watching the star-lit rim of the event horizon quickly recede farther and farther from view.

The rim gleams with a fierce strip of blinding light. Stars crest the steep fold in space-time before tipping over the edge and sinking into the same inkwell that our ship plunges into, drowning. The silver-and-gold trim of the hole in space thins the deeper we sink, casting our various skin hues in shades of deep violet.

"We're watching the end of the Universe up there," Lex Bethel remarks from where he leans against the bulkhead doors, corded arms crossed over his broad chest. He says it matter-of-factly, with no hint of wistfulness or fear. He's our engineer and has a reputation of steely nerves to maintain. If this



catastrophe were a little less catastrophic, he might lighten the mood with a joke. I would nod, Emma would smile. The shadow-frequency would hum. Naturally more dour, Julius and Myreen would do absolutely nothing at all.

But no joke follows. Instead, Myreen uses their raspy tenor to say, in a quartet of harmonized voices: "You can't be sure of that. Perception and reality are two very different things."

The shadow-frequency buzzes in the air around Myreen, predictably agreeing with its guardian. Myreen and the shadow-frequency work together seamlessly, in one voice. I would expect no less from a communications officer and their sentient radio.

"It doesn't matter either way. Not to us," Julius mutters, his hand tightening around Emma's as they face the scene outside the window, together... as if holding hands can stave off the doom of sinking into a black hole.

I say nothing. My people are introspective, to the point of losing our ability to speak verbally several generations ago — my mental voice is a by-product of conversations with the shadow frequency, who watched way too many late-twentieth-century sitcoms as a young, sentient radio. Like, way too many.

But the silence serves my people well, as we're natural data collectors and password/secret-keepers for half the fleet. Despite our child-like appearance,

we're also typically much older than our crewmates, with life spans counted in centuries, not decades. Lex Bethel, with his grey beard and wrinkled laugh lines, is closest to me in age, with only 67 years between us.

Our tendency to silence sells the whole ancient-child with superior knowledge of the Universe 'thing'.

But Emma suddenly breaks off her hand-holding with Julius and turns to me. She asks, "Is this the end, Lysa-child? Is there anything we can do? Can you search your memories to find out what our options are?"

If she thinks I haven't searched and researched the data cluttering up the secondary brain-space stored in my wrists and forearms, she's just being super stupid. Or hopeful. I suppose the line between those two constructs

is about as wide as that gold ring on the rim of this ridiculously deep gash in the cosmos.

I shake my head. But they all look so desperate and child-like in that moment, even Lex Bethel, that I decide I should say something. So I reach out and borrow the shadow-frequency, letting Myreen speak for me.

"The end of the Universe is an illusion. We'll pass through the other side and emerge in new beginnings," Myreen's many voices flicker over my words, both prophetic and mystical. The voices glimmer with unsubstantial hope and pure lies.

My lies soothe them, even if they don't soothe me. We're falling into a black hole. I feel swallowed up, guzzled down the throat of an apathetic god of cold starlight. The sides of this deep well close in. Darkness seeps through the cracks in our ship. If I were to stretch my hand out beyond the ship doors, I would feel the chill of an empty hole — and there's nothing to fill it. Not Myreen's voices or Lex Bethel's laughter. Not even the bickering of Emma and Julius, which, despite the whole 'falling in love' nonsense, suddenly picks up right where they left off.

"I told you to stay to the left of that cluster."

An eternity passes... in just about three minutes. ■

Gretchen Tessmer is an attorney/writer based in the US-Canadian borderlands of northern New York. Follow her on Twitter: @missginandtonic.

ILLUSTRATION BY JACEY



Cover art: Marco Melgrati

Editorial

Herb Brody,
Richard Hodson,
Lauren Gravitz,
Jenny Rooke,
Elizabeth Batty

Art & Design

Mohamed Ashour,
Wesley Fernandes,
Kate Duncan,
Denis Mallet

Production

Nick Bruni, Karl
Smart, Ian Pope

Sponsorship

Stephen Brown,
Claudia Danci

Marketing

Nicole Jackson

Project Manager

Rebecca Jones

Creative Director

Wojtek Urbanek

Publisher

Richard Hughes

Editorial Director

Stephen Pincock

Magazine Editor

Helen Pearson

Editor-in-Chief

Magdalena Skipper

People wear their health on their skin. As the body's largest organ, skin is our first line of defence against infection and injury; it is also crucial for temperature regulation and vitamin production, and its sensory capabilities help us to interact with the environment (see page S84). Skin is also very visible, with its appearance telegraphing vigour or disease — leaving those with certain skin conditions vulnerable to detrimental psychological effects. Therefore, although most skin diseases are not life-threatening, they are a leading cause of disability and researchers are working hard to find ways to help.

Understanding of the factors that affect skin health is improving steadily. Some of skin's stealthiest insults are being traced to the environment: ultraviolet radiation, air pollution and pesticides can be absorbed by skin, where they cause conditions that range from irritation to cancer (S89). A poor diet can also affect skin's health, increasing people's vulnerability to melanoma and other skin conditions (S94). And it is becoming clear that the bacteria, viruses and fungi that live on skin contribute to health, too: some might exacerbate certain conditions whereas others might offer protection (S91). For all of these factors, researchers have only started to translate their findings into specific, therapeutic advice.

Skin's full regenerative potential has yet to be unlocked, but many are searching for better ways of healing burns and deep wounds (S86). Materials scientists are creating electronic skins that could be useful for monitoring patients' vital signs or building improved prosthetic limbs (S96). And fresh treatments are on the horizon for people with vitiligo; however, a small-yet-vocal group who are advocating for acceptance rather than a cure could create a schism in the vitiligo community (S99).

We are pleased to acknowledge the financial support of Almirall in producing this Outlook. As always, *Nature* has sole responsibility for all editorial content.

Lauren Gravitz
Contributing editor

CONTENTS

S84 SKIN

Superpowered skin

The body's largest organ

S86 REGENERATION

The secrets of healing without scars

Potential treatments for burns and wounds

S89 ENVIRONMENT

When the first defence fails

Skin absorbs more pollutants than expected

S91 MICROBIOME

Community effort

Do microbes on skin protect or pose a threat?

S94 NUTRITION

Edible skin care

How certain diets could benefit skin

S96 ELECTRONICS

Beyond the biological

Incorporating a sense of touch into prostheses

S99 PERSPECTIVE

The eye of the beholder

Consider the needs of all members of the vitiligo community, says *John Harris*

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at go.nature.com/e4dwzw

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2018).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Skin* supplement can be found at www.nature.com/collections/skin-outlook. It features all newly commissioned content as well as a selection of relevant previously published material that is made freely

available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

Site licences (www.nature.com/libraries/site_licences): Americas, institutions@natureny.com; Asia-Pacific, <http://nature.asia/jp-contact>; Australia/New Zealand, nature@macmillan.com.au; Europe/ROW, institutions@nature.com; India, npgindia@nature.com. Personal subscriptions: UK/Europe/ROW, subscriptions@nature.com; USA/Canada/Latin America, subscriptions@us.nature.com; Japan, <http://nature.asia/jp-contact>; China, <http://nature.asia/china-subscribe>; Korea, www.natureasia.com/ko-kr/subscribe.

CUSTOMER SERVICES

Feedback@nature.com
Copyright © 2018 Springer Nature Ltd. All rights reserved.

SUPERPOWERED SKIN

The skin is the body's largest organ and has several, diverse functions. As well as being a physical barrier, it has immune and sensory properties.
By Julie Gould; illustration by Lucy Reading-Ikkanda

UNDER THE SURFACE

Skin's most important role is to protect the body from the environment. It comprises three main layers: the epidermis, the dermis and subcutaneous fat. Most of the body is covered in hairy skin but the palms of the hands and the soles of the feet are covered in hair-free (glabrous) skin.

Epidermis

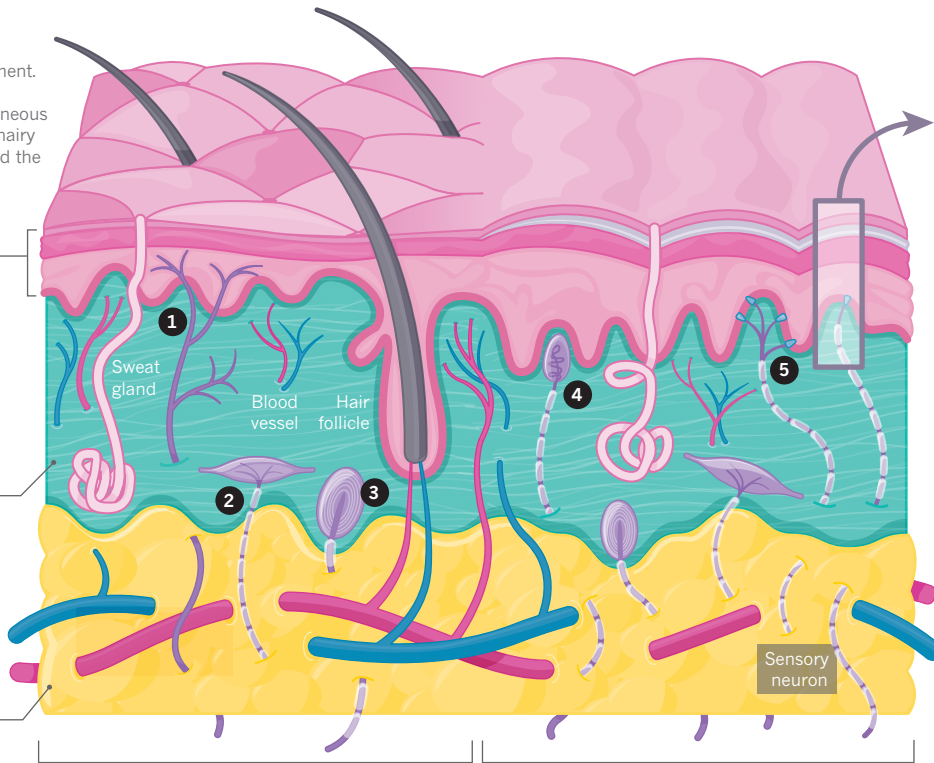
The outermost layer of skin acts as a mechanical and antimicrobial barrier, and consists of several layers. Its top part, the stratum corneum, prevents water from leaving the body and toxic substances from entering.

Dermis

Nerve endings in skin's middle layer help people to feel sensations such as itching, pain, pleasure and heat. The dermis produces sweat and oils, and contains hair follicles. It also hosts a variety of immune cells.

Subcutaneous fat

Skin's deepest layer is sandwiched between the dermis and skeletal muscles. Its roles include fat storage, connecting the dermis to muscle and bone, and controlling body temperature.



Hairy skin

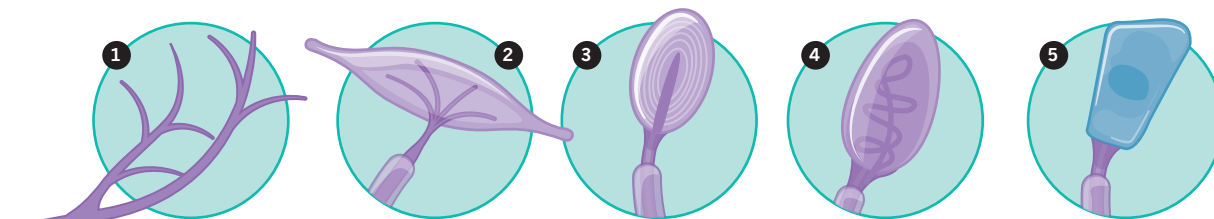
More than 90% of the body is covered by hairy skin¹. It is involved in perceiving a variety of tactile sensations, including those that form part of social exchanges, and the ability to detect the presence of foreign objects. In hairy skin, the epidermis is less than 0.1 millimetres thick and the dermis is 1–2 millimetres deep.

Glabrous skin

Hair-free skin is found mainly on the palms and soles. It is innervated by specialized nerves that help us to understand subtle tactile details. Such skin is thicker than hairy skin; the epidermis is about 1.5 millimetres thick and the dermis is about 3 millimetres deep.

SENSATIONAL SENSITIVITY

Skin's somatosensory system comprises more than a dozen subtypes of sensory neuron, but only those involved in tactile sensation are well understood. Such neurons enable skin to react to and interpret myriad stimuli, including temperature gradients, pressure and physical damage.



C fibre

These unmyelinated nerve fibres are found only in hairy skin. Although sensitive to indentation, they are most active when a stimulus moves slowly across the skin's surface.

Ruffini ending

Found in the dermis of both hairy and glabrous skin, these sensory receptors respond optimally to stretching of skin.

Pacinian corpuscle

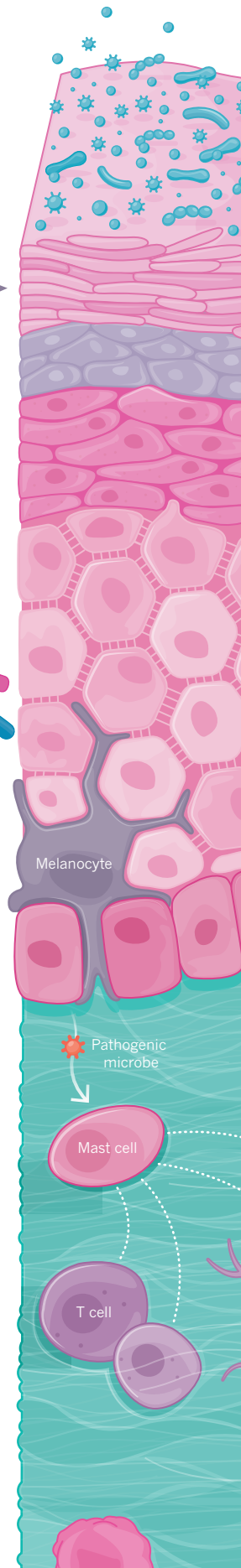
Located deep in the dermis of both types of skin, Pacinian corpuscles respond to high-frequency vibration.

Meissner corpuscle

These nerve receptors lie just beneath the epidermis of glabrous skin, where they detect movement across the skin and fluttering touch.

Merkel cell

Part of the stratum basale of the epidermis, these cells help to relay information about the texture, curvature and shape of objects. Merkel cells are most dense in glabrous skin.



PROTECTIVE LAYERS

Skin's epidermis and dermis help to protect the body from microbes, pollutants, ultraviolet radiation and excessive loss or absorption of water.

MICROBIOTA

The harsh environment of skin's surface — dry, nutrient-poor and acidic — houses a mixture of bacteria, fungi and viruses that help to fend off invading species and communicate with immune cells.

EPIDERMIS

Five sublayers continuously rebuild the skin's surface.

STRATUM CORNEUM

A layer of dead, flattened cells filled with the protein keratin that protects the body from friction and water loss.

STRATUM LUCIDUM

Also composed of dead cells, this layer is found only in glabrous skin and is packed with lipid-rich eleiden, which helps to keep water out.

STRATUM GRANULOSUM

A layer made mostly of mature keratinocytes that migrate from the stratum spinosum. It also helps to waterproof skin.

STRATUM SPINOSUM

Keratinocytes — mature basal cells — produce keratin, which comprises the basic structure of skin. Immune cells called Langerhans cells that inform the immune system about invading microbes are also present.

STRATUM BASALE

The deepest layer of the epidermis contains continually dividing basal cells, which push older cells upwards. It also contains melanocytes, which control skin pigmentation. When melanocyte DNA is damaged by ultraviolet radiation, any resulting uncontrolled cell growth can lead to the skin cancer melanoma.

30–45 DAYS

Time it takes for basal cells to mature and migrate to the top of the epidermis².

DERMIS

The dermis provides skin with support and elasticity through the proteins collagen and elastin. It also offers protection from pathogenic microbes and toxic substances. Four types of immune cell contribute to this line of defence.

Mast cell

A cell that identifies pathogenic species and then releases chemical signals to attract other immune cells.

T cell

A white blood cell that remembers microbes encountered previously.

Dendritic cell

Presents parts of pathogenic microbes to other immune cells.

Macrophage

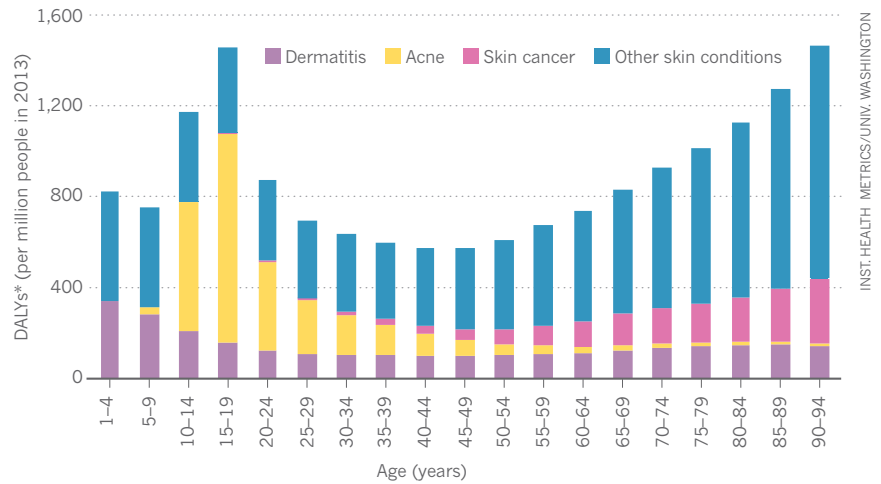
Helps to clear cellular debris.

BARRIER BREAKDOWN

Despite its many superpowers, the skin is not infallible. Because of its visibility, diseases that affect the skin can have psychological as well as physical effects.

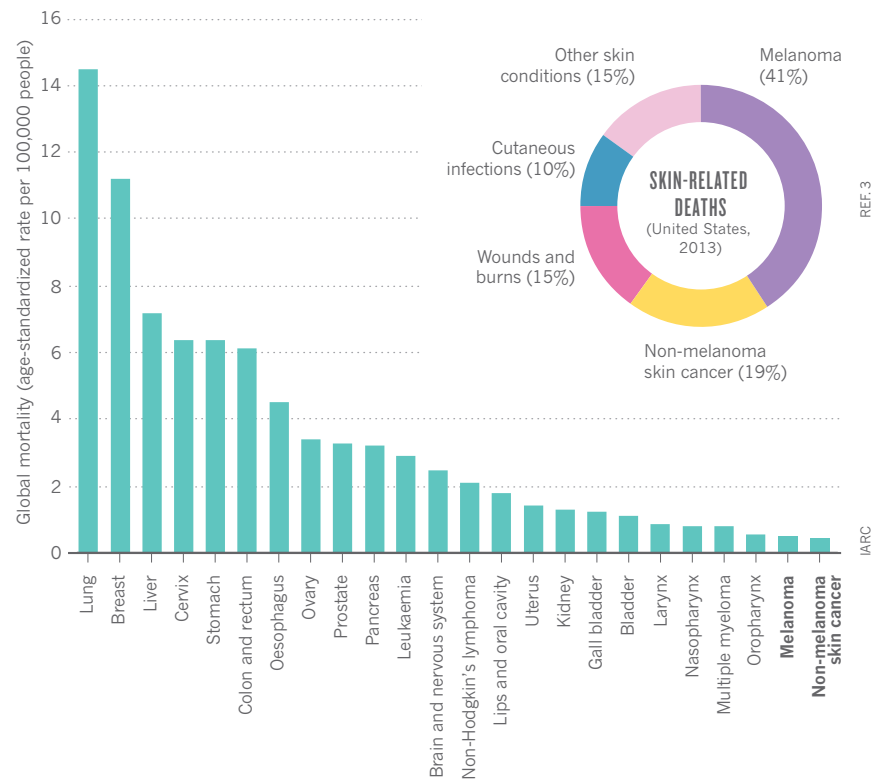
BURDENSOME BOUNDARY

Skin disease's worldwide burden can be quantified in terms of the disability-adjusted life year (DALY), which reflects a lost year of healthy life. The main burden falls on people aged 15–19, mostly owing to acne vulgaris. From the age of 50, there is a slow increase in burden as skin loses function and the incidence of skin cancer rises. Dermatitis, including eczema, persists throughout life; people tend not to outgrow the condition but learn to better manage it.



CANCER COMPARISON

Melanoma kills more people worldwide than does non-melanoma skin cancer, even though non-melanoma is much more common. However, deaths from melanoma are dwarfed by those from other cancers. Skin cancer was responsible for about 60% of US skin-related deaths in 2013.



Sources: 1. Zimmerman, A., Bai, L. & Ginty, D. D. **346**, 950–954 (2014). 2. US National Cancer Institute. *Layers of the skin* <https://training.seer.cancer.gov/melanoma/anatomy/layers.html> (US National Institutes of Health, 2018). 3. Lim, H. W. et al. *J. Am. Acad. Dermatol.* **76**, 958–972 (2017).



REGENERATION

The secrets of healing without scars

Skin regeneration is an imperfect process that is impeded by a host of factors. Working out the part played by each could lead to fresh approaches to treating burns and scars.

BY CASSANDRA WILLYARD

The body's largest organ might seem barely more than cellular wrapping paper, but skin has roles that range from fending off microorganisms to regulating body temperature. It also has a considerable flaw: severely damaged skin can heal, but it can't regenerate. Instead, it forms scars. These marks are not just cosmetic defects. Scar tissue can inhibit a person's movement and, because it lacks sweat glands, prevent the body from cooling off. Although scars seem to be thicker than normal skin, the tissue is actually weaker.

Scarring seems to be an inevitable part of being human. But three decades ago, it became clear that the youngest patients don't scar.

When Michael Harrison, a paediatric surgeon at the University of California, San Francisco, began to perform the first ever surgeries on fetuses, he noticed something curious about the babies who survived. Incisions he had made in them in the womb seemed to heal without scarring.

Harrison asked Michael Longaker, a postdoctoral researcher in his laboratory, to investigate the phenomenon. Longaker was sceptical. Because his boss was the only physician who was performing fetal surgeries, he says, "My first reaction was, 'Gosh, that doesn't seem like a big health-care problem because you're the only one making [fetal] wounds.'" But it didn't take long for Longaker to understand the potential implications: by

deciphering what drives this *in utero* healing, he might discover ways to prompt scar-free healing outside the womb. "My reluctant one year in the lab became four," Longaker says. "I became obsessed with scarring."

Longaker, now a plastic surgeon with a focus on regenerative medicine at Stanford University in California, has not yet unravelled the mystery completely. Nor have other researchers. Although many studies have provided valuable insight into how scarring occurs, they have yielded few clinically useful treatments. "There's been some improvement," says Stephen Badylak, deputy director of the McGowan Institute for Regenerative Medicine at the University of Pittsburgh in Pennsylvania. But it's still far from the expectations raised by

MARCO MELGRATTI

the hype of the work that began in the 1980s.

Yet many researchers are cautiously optimistic that a better understanding of the mechanisms that lead to scarring will pave the way for innovative strategies for reducing the formation of scar tissue. In September, the US Food and Drug Administration approved the first treatment to involve a 'spray-on' skin, and numerous other skin-healing products are in clinical trials. The field of skin regeneration is moving in a different direction, Badyalak says. Rather than growing skin in Petri dishes in the lab, and then transplanting it onto people, researchers are using the body as a bioreactor and encouraging skin to do what it did during fetal development — regenerate. They want to find out more about how scarring occurs, as well as how it might be stopped.

EVOLUTIONARY ADVANTAGE

Cut the skin and it will bleed. And then it will heal. Initially, a clot forms to staunch blood flow, which kicks off a massive inflammatory response. Immune cells flood the region to clear bacteria and debris, while cells called keratinocytes in skin's outer layer divide rapidly in a race to close the wound and prevent infection. Next, the wound begins to fill. Spindle-shaped cells known as fibroblasts migrate to the damaged area and churn out collagen and other proteins that provide tissue with structure. Within three weeks of the injury occurring, the wound has healed.

But such speedy healing has a major downside. These quick repairs often result in scars, particularly when the wound is deep. In healthy skin, collagen fibres form a lattice. But during wound healing, fibroblasts lay down collagen fibres parallel to each other, which creates tissue that is stiff and weak. That's because evolution has selected speed over perfection: before the discovery of antibiotics, slow healing would probably have meant acquiring an infection or experiencing prolonged bleeding. "It's really a matter of survival versus aesthetics," says Jeff Biernaskie, a stem-cell biologist at the University of Calgary in Alberta, Canada.

When such repairs to skin are small, they don't pose much of a problem. But large scars can be life-changing. Scar tissue "doesn't have the stretch and the mobility and the range of motion that normal skin does," says Angela Gibson, a burn surgeon who studies wound healing at the University of Wisconsin School of Medicine and Public Health in Madison. That can be especially problematic when scars cover joints. Imagine, Gibson says, not being able to hold a fork or to raise your arms to wash your hair.

But scarring might not be inevitable. Fetal skin begins to scar only late in gestation, which suggests that human skin possesses at least some regenerative capabilities. All researchers have to do is to work out how to unlock them.

FANTASTIC FIBROBLASTS

Fetal wounds are not the only wounds that are resistant to scarring. Thomas Leung,

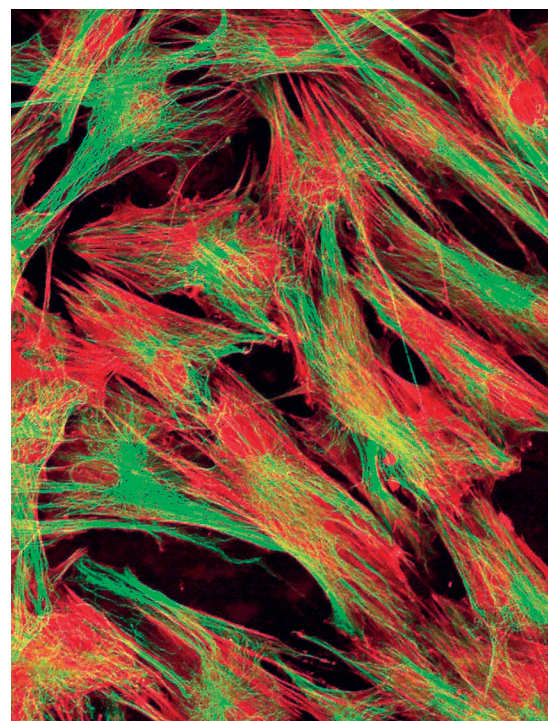
a dermatologist at the Perelman School of Medicine at the University of Pennsylvania in Philadelphia, noticed that older people often develop thinner scars than do younger adults. To understand why, Leung turned to mice. He and his colleagues compared wound healing in young and old mice by punching holes in the rodents' ears¹. In one-month-old animals, such wounds healed with a thick scar and never closed fully — similar to earring holes in people, Leung says. In 18-month-old mice, which are roughly equivalent to 65-year-old people, healing took longer, but the holes closed completely, and with less scarring. The same observations held for wounds on the backs of the mice.

Leung and his colleagues wondered whether a component of the blood of young mice promotes scar formation. To test the idea, they joined together old and young mice, giving them a shared circulatory system through a surgical technique called parabiosis. The team found that exposure to the blood of young animals caused wounds in elderly mice to scar¹. Further experiments revealed the probable culprit: *Cxcl12*, a gene that encodes a protein called stromal cell-derived factor 1 (SDF1). When the team knocked out SDF1, even wounds in young animals healed with minimal scarring. This discovery suggests a route towards scar-free wound healing in people: suppressing the activity of *CXCL12*.

In fact, there's already a drug on the market that interferes with the SDF1 pathway — plerixafor. The drug is used to mobilize stem cells from bone marrow in people with certain types of cancer. Leung and his colleagues hope to test whether plerixafor can minimize the recurrence of keloids — thick, raised scars that tend to keep growing — in a clinical trial. The team is also looking at how SDF1 promotes initial scar formation.

Scarring is a complex process, and SDF1 is only part of the story. Fibroblasts are another prominent player. These cells have long been blamed for scar tissue. "We've had this assumption that fibroblasts are all the same," Biernaskie says. But research in the past five years has revealed that fibroblasts comprise a diverse group of cells, and that some seem to have a larger role in scar formation than do others.

In 2015, Longaker and his colleagues conducted an inventory of the fibroblasts on the skin of a mouse's back². When they created a wound on the back, they found that only one of two lineages of fibroblast — expressing homeobox protein engrailed-1 — was responsible for the formation of most scar tissue. And when the team disabled those cells in mice, wounds healed more slowly but also formed less scar tissue, similar to what happened in mice that lack SDF1. Longaker thinks that if he and



Fluorescence micrograph of human skin fibroblasts.

other researchers can find a way to identify and block the same fibroblasts in people, it might be possible to prompt wound healing to follow a more regenerative pathway. "I would be disappointed if we're not doing something like that in humans in the next five to seven years," he says.

Although some fibroblasts are clear drivers of scar formation, other research suggests that fibroblasts also contribute to regenerative healing. About a decade ago, George Cotsarelis, a dermatologist at the Perelman School of Medicine, and his colleagues were trying to develop a mouse model to understand the role of stem cells in hair follicles. Scientists had long thought that when an adult hair follicle is lost, it is gone for ever. But then the team noticed something odd: when they made a large wound on the back of a genetically normal mouse, hair regrew in the middle of the wound³.

Even more strangely, skin around hair follicles seemed to be normal, and a layer of fat formed beneath — something that doesn't usually occur under scar tissue. In 2017, a team led by Cotsarelis showed in mice that new hair follicles secrete growth factors called bone morphogenetic proteins (BMPs) that can transform fibroblasts into fat cells⁴. "The really cool part," Costarelis says, is that "once you get a hair follicle, it kind of normalizes the skin".

Human fibroblasts also seem able to make the leap from fibroblast to fat. When the team took such cells from a keloid scar and exposed them to a BMP, or placed them near a BMP-secreting hair follicle, they too turned into fat cells. These findings suggest that it might be possible to prod injured skin towards regeneration rather than scar formation. But translating the work into a treatment protocol poses considerable difficulties, Cotsarelis says. Skin

"The really cool part is that once you get a hair follicle, it kind of normalizes the skin."



Reindeer antler velvet has regenerative properties.

regeneration will require the right signals to be delivered at the right time, and at the right dose. For example, “When hair follicles form, their spacing is determined by gradients of growth factors,” he says. Altering those gradients, even slightly, might alter the follicle pattern or even function. “Precision is really required,” he says.

A MORE PERFECT MODEL

The mice in which most research on wound healing is performed differ from people in important ways. Their skin is loose, whereas that of humans is tight. Furthermore, mouse wounds heal by contraction: such wounds pull together rather than filling in. “I don’t know how you can even begin to think you could test something there and then translate it to humans,” Gibson says.

In search of a better model, in 2009, Ashley Seifert, a developmental and regenerative biologist at the University of Kentucky in Lexington, travelled to Kenya and began to study African spiny mice (*Acomys kempfi* and *Acomys percivali*) — species with a unique defence mechanism. Because their skin tears easily, these mice can escape the jaws of predators. Seifert expected to find that such mice had speedy wound-repair processes or ways of preventing infection. But what he and his colleagues found was much more intriguing: spiny mouse wounds heal relatively scar free⁵.

The spiny mouse is one of only a few mammalian models of skin regeneration. But such mice provide a comparative framework. Seifert can punch a hole in the ear of a spiny

mouse, which regenerates, and another in the ear of a conventional lab mouse, which does not, and then evaluate how the healing process differs. His team is now beginning to define those differences.

Some seem to involve the immune system. Researchers tend to view inflammation as an impediment to regenerative healing. Accordingly, the difference between scar formation in adults and the fetus might be that adults mount a strong inflammatory response after injury whereas a fetus does not. But a connection between inflammation and regeneration has been difficult to establish. Efforts to prevent scar formation by suppressing inflammation haven’t panned out, Seifert says. And he and his colleagues have found, at least in spiny mice, that inflammation does not preclude regenerative healing. In the wild, these mice mount a strong inflammatory response yet still manage to regenerate skin.

“We know that too much inflammation is bad. And we know that no inflammation isn’t helpful either,” Seifert says. In 2017, he and his colleagues showed that macrophages, immune cells that are a key orchestrator of inflammation that is typically associated with scarring, are also required for regenerative healing in spiny mice⁶. Now, the team is trying to determine which factors might tip macrophages and other immune cells away from scarring pathways and towards regeneration.

A much larger mammal — reindeer (*Rangifer tarandus*) — is also providing insight into the regenerative potential of skin. Both male and female animals sprout new antlers each year. The downy velvet that covers the antlers as they grow is remarkably similar to human skin — thick with blood vessels, hair follicles and sebaceous glands. But it differs in one important way. “If we wound the velvet, it regenerates perfectly,” Biernaskie says. “It’s really a beautiful and powerful model for skin healing.”

That capacity for regeneration seems to be inherent to the velvet. Biernaskie and his colleagues are now comparing changes in gene expression during wound healing in two anatomical areas of reindeer — skin on their backs, which doesn’t regenerate, and antler velvet, which does. They hope that the comparison will help them to better understand the signals that prompt velvet to regenerate, and perhaps lead them to treatments that promote regeneration and prevent scarring. “We could start to develop cocktails of drugs where we could mimic those signals,” Biernaskie says.

They hope that the comparison will help them to better understand the signals that prompt velvet to regenerate, and perhaps lead them to treatments that promote regeneration and prevent scarring. “We could start to develop cocktails of drugs where we could mimic those signals,” Biernaskie says.

BENCH TO BEDSIDE

Skin regeneration is still a distant goal, but several companies are working to bring wound-healing therapies to market. The

spray-on skin system approved by the Food and Drug Administration earlier this year, and marketed as ReCell by biotechnology company Avita Medical in Valencia, California, is an example of an early success.

To prepare the treatment, surgeons remove a piece of skin about the size of a postage stamp from the patient and douse it with an enzyme that liberates skin’s component cells: fibroblasts, keratinocytes and pigment-producing melanocytes. These cells are then loaded into a nozzled syringe and sprayed onto the patient’s wound. People with burns who require skin grafts typically receive pieces of skin that are harvested from unaffected parts of their bodies. Surgeons take only the top layers of skin to create these grafts, which are known as split-thickness grafts. One clinical trial showed that in people with second-degree burns, which affect both skin’s epidermal and dermal layers, the ReCell therapy works as well as do conventional grafts, but requires much less donor skin⁷. Although split-thickness grafts can be cut into a mesh that covers an area about three times their size, ReCell can treat skin wounds that are 80 times larger than the donor piece of skin. ReCell can also be combined with meshed grafts to treat deeper burns.

Gibson is testing an alternative treatment for burns, a skin substitute called StrataGraft. It comprises two layers of collagen: a bottom layer that is seeded with human fibroblasts and a top layer that is seeded with cells that give rise to keratinocytes. The therapy originated at the University of Wisconsin, but is now being developed by Mallinckrodt Pharmaceuticals in Staines-upon-Thames, UK. One of the first clinical trials of StrataGraft, published in 2011, showed that it did not induce an acute immune response⁸, and the substitute is now being tested in a phase III trial.

Such therapies could be a boon for people with burns. Other companies are working on treatments for tricky-to-heal wounds, such as ulcers in people with diabetes or bedsores. “The market size is just gigantic,” Badylak says. But the main goal of these treatments is to promote better healing, rather than to prompt skin to regenerate. Achieving that next step — scar-free healing — is “a tall order to fill,” Gibson says. However, she is optimistic that if clinicians who treat skin wounds collaborate closely with researchers who are working to understand scarring, the problem can be solved. “That’s when the science will move forward,” she says. ■

Cassandra Willyard is a science journalist in Madison, Wisconsin.

1. Nishiguchi, M. A., Spencer, C. A., Leung, D. H. & Leung, T. H. *Cell Rep.* **24**, 3383–3392 (2018).
2. Rinkevich, Y. et al. *Science* **348**, aaa2151 (2015).
3. Ito, M. et al. *Nature* **447**, 316–320 (2007).
4. Plikus, M. V. et al. *Science* **355**, 748–752 (2017).
5. Seifert, A. W. et al. *Nature* **489**, 561–565 (2012).
6. Simkin, J., Gawriluk, T. R., Gensel, J. C. & Seifert, A. W. *eLife* **6**, e24623 (2017).
7. Holmes, J. H. IV et al. *J. Burn Care Res.* **39**, 694–702 (2018).
8. Centanni, J. M. et al. *Ann. Surg.* **253**, 672–683 (2011).



The Ruhr Valley is an industrialized region of northwest Germany.

ENVIRONMENT

When the first defence fails

Pollutants in the environment cause more harm to skin than once thought, with effects that can range from irritating to deadly.

BY ELIZABETH SVOBODA

In the hazy streets of some of Asia's largest cities, people often wear face masks to help avoid inhaling airborne pollutants into their lungs. More than a decade ago, dermatologist Jean Krutmann began to wonder whether such pollutants also affected the body's largest organ, the skin. He and his colleagues began to study people from Europe and Asia who were exposed regularly to vehicle exhaust emissions such as the gas nitrogen dioxide or particulate matter, tracking changes in their health over time.

Krutmann's initial results were traffic-stopping. People who are exposed to common air pollutants have higher rates of chronic skin inflammation and more age spots than do those who live in cleaner areas¹. "I was the one who bet nothing would come out of this," says Krutmann, director of the Leibniz Research Institute for Environmental Medicine in Düsseldorf, Germany. "We were all surprised to see that there was a strong association."

As the soft, flexible barrier that surrounds the body's tissues, skin is porous enough to soak up moisture, absorb medications from adhesive patches and release protective oils. But the same property also makes

skin vulnerable to assault by chemicals in the environment. And ultraviolet radiation in sunlight can cause premature skin ageing and skin cancer such as squamous cell carcinoma, a mechanism that has been known about for several decades. More recently, research has broadened dramatically to reveal the serious harm that can be inflicted on skin by air pollutants, pesticides and other common chemicals.

"Just in the last ten years, there's so much science about new environmental stressors," says Whitney Bowe, a dermatologist at Icahn School of Medicine at Mount Sinai in New York City. And although damage from such exposure is usually confined to skin, it can also go much further. Pollutants and chemicals that pass through skin can contribute to conditions such as asthma or breast cancer, so researchers are investigating fresh ways to keep dermal incursions at bay.

MAKING AN ENTRANCE

Skin can be affected by the environment in a variety of ways. The most obvious is through direct exposure to chemicals: submerging an arm in a vat of acetone or benzene would enable the substance to enter the skin by diffusion.

Ultraviolet radiation does not penetrate skin in the same way, but can trigger a destructive

chemical reaction. When it strikes molecules on the skin's surface that contain oxygen, unstable compounds called free radicals are created. To stabilize themselves, free radicals steal electrons from nearby molecules — a process known as oxidation. This can damage the DNA of skin cells, which leads to tissue inflammation, accelerates skin ageing and promotes mutations that contribute to cancer.

But chemicals in the environment often enter skin in more insidious ways. "Skin exposures are stealthy," says physical scientist Frederick Frasch, a coordinator at the US National Institute for Occupational Health and Safety in Morgantown, West Virginia. "If a chemical is toxic through ingestion or inhalation, it will also be toxic through skin absorption." The smogs of Beijing, the pesticide hazes of the Central Valley in California and the smoke plumes that rise from wildfires in the western United States all contain an array of hazardous chemicals — many of which are still being identified.

Many airborne pollutants are so small that they enter pores in the skin like pebbles dropping into a cup, Bowe says. Others, including the polycyclic aromatic hydrocarbons found in vehicle exhaust emissions or wildfire smoke,

LUKAS SCHULZE/GETTY



Study participants are exposed to airborne phthalates through the skin.

are ‘fat-loving’ (lipophilic) and can easily pass through the fat-filled spaces between skin cells. They then enter the circulatory system, where they can cause widespread effects.

The most common problem associated with exposure to environmental pollutants is localized skin irritation, says Sean Semple, an occupational-health scientist at the University of Stirling, UK. But a number of pollutants can cause serious, long-term issues. High concentrations of air pollutants have been linked to infertility, asthma and even some cancers. Pesticides can impair brain and nerve function over time. And phthalates — chemicals that are used to make plastic more flexible and are released when it degrades — have been connected to hormonal imbalances in children and abnormal reproductive development in fetuses.

MAPPING THE EXPOSURE LANDSCAPE

One of the main goals of environmental scientists is to document the scope and severity of air pollution’s effects on skin health. “Air pollution is not [only] in megacities in East Asia — it’s clearly in Western countries,” Krutmann says. “You cannot avoid it.” He has been following a cohort of older women in the Ruhr Valley, an urban region of Germany, who are exposed to road-traffic pollution at levels similar to those in countless other communities worldwide. After adjusting for socio-economic status, smoking habits and exposure to ultraviolet radiation, among other factors, Krutmann concluded that such pollution contributed to the accelerated skin ageing he saw in study participants¹. He found similar results in a cohort of Han Chinese women from an urban area. And earlier this year, Krutmann and his colleagues built on that research to show that older women who are exposed to traffic pollution have increased rates of eczema, a skin condition characterized by inflammation and scaly red rashes².

In the laboratory, researchers are working to understand exactly why air pollution has these

effects. A 2017 study that included scientists at Guangdong Environmental Monitoring Center in Guangzhou, China, confirmed that exposing immortalized human skin cells to particulate matter — a stew of microscopic dust, soot, exhaust and smoke particles usually suspended in air — leads to the formation of free radicals, DNA damage and cell death³.

Skin is also vulnerable to airborne phthalates. Researchers have long known that these hormone-disrupting compounds are dangerous when ingested. But Charles Weschler, who investigates pollutant exposure at Rutgers University in Piscataway, New Jersey, suspected that phthalates might also pose a danger when they leach from household goods into the home environment. To measure how efficiently airborne phthalates are absorbed by skin, Weschler confined a small group of people to rooms filled with air containing elevated levels of phthalates for six hours. In one trial, participants donned a breathing hood so that they inhaled filtered air; in another, they had no hood and were asked to breathe normally.

Weschler found that participants absorbed about the same amount of phthalates — and sometimes even more — through their skin as they did through their lungs⁴. And he thinks that the level of phthalate absorption by skin might be even more dramatic in real-world settings such as the home, where chemicals accumulate over long periods. “The modelling suggested that if we had kept those experiments going for two days, the uptake would have been five times greater,” he says.

Pesticides used by farmers and gardeners might pose a similar threat to skin. Earlier this year, researchers at Griffith University in Nathan, Australia, published a study in which they monitored farm workers who were applying the common pesticide chlorpyrifos to rice paddies in Ghana. The team measured how much pesticide residue passed through the workers’ clothing to contact their skin. From those data, they calculated that many of the

workers absorbed a dose of the pesticide that was several times higher than the dose known to increase risk of developing acute adverse health effects such as confusion and intestinal distress⁵. Pesticides have long been linked to a variety of skin conditions, including contact dermatitis, acne and even melanoma. The results from Ghana suggest that these disorders could be a direct result of pesticide absorption by skin.

But there is still much that researchers do not understand about the levels of skin exposure to pollutants throughout people’s lives. Laura Vandenberg, an environmental-health scientist at the University of Massachusetts Amherst, studies the effects of a common chemical called bisphenol A on skin. In one study, she investigated the exposure of skin to bisphenol A from thermal paper used in shop receipts⁶. On the basis of this and other experiments, Vandenberg thinks that some of the agencies responsible for chemical risk assessments, such as the European Food Safety Authority, probably underestimate people’s exposure through skin to various chemicals. But assessing precise levels of exposure, and predicting effects on health, can be exceedingly difficult. For example, two people who do different jobs at the same farm might receive dramatically different levels of pesticide exposure through skin. And genetic variation makes it even harder to predict the specific effects of pesticide exposure on the skin and general health of individuals.

The best way to assess the connection between skin absorption, levels of environmental exposure and genetic contributions, as well as their combined effects on health, is to conduct large-scale studies. But for now, the most prudent thing that people can do is to limit their exposure to pollutants. This includes taking common-sense precautions such as applying high-sun-protection-factor sunscreen to exposed skin and wearing protective clothing.

Another way to protect skin, particularly from free radicals, Krutmann says, is to use creams that are rich in antioxidant compounds. Such creams neutralize free radicals at the skin’s surface, helping to halt the cell-destroying cascade. Krutmann says that many of these products — especially those containing vitamin C or vitamin E — work well to limit damage in cells. Skin’s absorption of environmental pollutants is a complex problem. But the solution, Krutmann thinks, might be as simple as bolstering the protective barrier that the skin already provides. ■

Elizabeth Svoboda is a science journalist in San Jose, California.

1. Hüls, A. *et al.* *J. Invest. Dermatol.* **136**, 1053–1056 (2016).
2. Schnass, W. *et al.* *Int. J. Hyg. Environ. Health.* **221**, 861–867 (2018).
3. Hu, R. *et al.* *Chin. Med. J.* **130**, 2205–2214 (2017).
4. Weschler, C. J. *et al.* *Environ. Health Perspect.* **123**, 928–934 (2015).
5. Atabilla, A. *et al.* *Chemosphere* **203**, 83–89 (2018).
6. Bernier, M. R. & Vandenberg, L. N. *PLoS ONE* **12**, e0178449 (2017).



Jack Gilbert investigates the movement of microbes between skin and the environment.

MICROBIOME

Community effort

Each person's skin carries a unique population of microbes that might help to protect skin, or increase its vulnerability.

BY EMILY SOHN

In December 2012, two months before the opening of a new, ten-storey hospital building in Chicago, Illinois, Jack Gilbert and his team descended with cotton swabs. In ten seemingly empty hospital rooms and at two nurse stations, they took samples from bed rails, floors and other surfaces to test for microbial life that had already moved in. After the building opened, the team returned

again and again, taking samples — including from hands, armpits and other body parts of patients — as often as once a day.

After a year of sampling, Gilbert, a microbiologist at the University of Chicago, had enough information to build up a picture of how microorganisms move between the environment and people's skin in a hospital setting. As soon as patients arrived, communities of microbes on the hospital surfaces began to reflect the groups that tend to live on skin,

including species of bacteria from the genera *Staphylococcus* and *Streptococcus*. But that wasn't surprising to Gilbert. Each hour, people shed around 37 million bacteria and 7 million fungi into the air.

More intriguing was how the microbial communities changed as patients came and went. Each person's skin carries a unique combination of microbes, known as a microbiota, the collective genomes of which are called a microbiome. During the first day of their stay, patients picked up microbes that had been left behind by the room's previous occupant. Soon, however, their own microbes took over¹. "Within 24 hours, the old patient's microbiome signature in the room completely disappeared," Gilbert says. "And it completely disappeared on the new patient's skin."

These findings add to a growing understanding of how communities of bacteria, fungi and viruses form on skin, and their potential for affecting health — particularly in hospitals, where drug-resistant bacteria are of increasing concern. In the past decade, researchers have characterized the types of microbe that thrive on skin and have investigated how those microbes colonize people early in life. They have also linked specific species of microbe to infections and skin conditions such as eczema and acne vulgaris — connections that could be more than simple associations, especially as skin-dwelling microbes might both cause skin disorders and prevent them.

Such tantalizing discoveries conjure up ideas of a fresh generation of treatments that improve health by adjusting the skin microbiota. Although the research is less than conclusive, both established and start-up cosmetics companies are enlisting scientific advisers to develop microbe-specific products to improve health and beauty. Many skin researchers, however, urge caution. "Long-term, I do see promise," says Julie Segre, a geneticist at the US National Human Genome Research Institute in Bethesda, Maryland. But, she adds, "We're really at the beginning stages."

TAKING ATTENDANCE

Millions of microbes cover the surface of skin. But studies of the skin microbiota have lagged behind research on gut microbes, Segre says. The reason is partly historical: initial research on microbiotas focused on the gut, which hosts the body's largest community of microbes — with potential implications for nutrition, digestion and immunity.

But as the body's first line of defence against pathogenic agents, skin is also an important gateway to the immune system. Gradually, research on the skin microbiota is catching up. So far, the bulk has focused on working out which microbes are present and how microbial communities form. Using next-generation genome sequencing, scientists have shown that a couple of hundred species of microbes, belonging to several main genera, thrive on the skin's surface — the epidermis. Microbes

have also been found to reside in the dermis, a deep layer of skin — a finding that challenges a long-held assumption that the dermis is sterile and protected from bacteria.

The make-up of the body's microbial communities varies by body part. Researchers have defined four basic environments using factors such as pH, moisture levels and temperature. Species of the genus *Propionibacterium* dominate oily sites such as the forehead. Moist sites such as elbow creases have fewer *Propionibacterium* and relatively more species belonging to the genus *Staphylococcus*. Feet are dominated by *Staphylococcus*. Fungi, mainly of the genus *Malassezia*, live all over the body but are most common in oily areas such as the face and back. Hair follicles are especially resource-rich environments for fungi and bacteria.

Such regional variations are consistent between people but, as Gilbert confirmed in his hospital study, each person's skin microbiota is unique enough to be identified from a swab taken from any part of their skin¹. Like that of the gut microbiota, the composition of the skin microbiota is remarkably consistent over time. For skin, this is probably because microbes that live in the dermis replenish the surface population as skin flakes off.

Over the course of about two years, Segre and her colleagues repeatedly sampled skin microbiotas at 17 sites on the bodies of 12 people². The team's findings showed that there were plenty of transient species of microbe, especially on the feet. But the dominant types tended to stay the same. "Even though our skin is the most exposed organ in our body — we're constantly bathing it, applying creams to it, touching things, being exposed to different people and environments — it's largely stable over time," says Julia Oh, a microbial geneticist at the Jackson Laboratory in Farmington, Connecticut, who worked with Segre on the study.

The microbiota's consistency has implications for drug development. When Oh and her collaborators bathed mice in "tons and tons of human skin microbes" three times a week for 30 weeks, they found that there was little colonization of the animals' skin. But as soon as a mouse received a cut to the skin, the microbes from human skin took over. These results are yet to be published and Oh acknowledges that the outcomes of transferring microbes from mouse to mouse, or from person to person, could differ. However, if a breach in the epidermis is necessary to alter the skin microbiota, Oh says, future therapeutics might need to agitate skin to have an effect.

The initial few months and years of a person's life seem to be a crucial time for the skin microbiota. In one of the first studies to scrutinize the interaction between the skin microbiota and the immune system in infancy, Tiffany Scharschmidt, a dermatologist at the University of California, San Francisco, found that adult mice were tolerant to *Staphylococcus epidermidis*, a skin bacterium found commonly in people, when the microbes were allowed to colonize their skin soon after birth³. The immune systems of the mice did not react when *S. epidermidis* was applied to skin abrasions. Although not usually pathogenic, this bacterium can sometimes cause infections.

But when mice encountered *S. epidermidis* for the first time as adults, their immune systems mounted an inflammatory response, which can impede wound healing, Scharschmidt says. This suggests that there is a period in early life in which it is beneficial for organisms to be exposed to the microbes that they will continue to encounter. In people, the first few years might

therefore be a key time for intervention, either by adding important microbes to the skin or by creating an environment that favours ideal microbiotas. "It speaks to a window of opportunity for establishing a healthy relationship with skin bacteria," Scharschmidt says. During that crucial time, she notes, "We don't want to limit exposure to healthy microbes."

MICROBES FOR ECZEMA

In 2009, around the time that researchers began to decipher the ecology of the skin microbiota, immunologist and dermatologist Richard Gallo and his team at the University of California, San Diego, in La Jolla, published a landmark study that hinted at potential treatments for inflammatory skin conditions. *S. epidermidis*, they found, produces a substance that suppresses inflammation — both in human skin cells in the laboratory and in the skin of mice⁴.

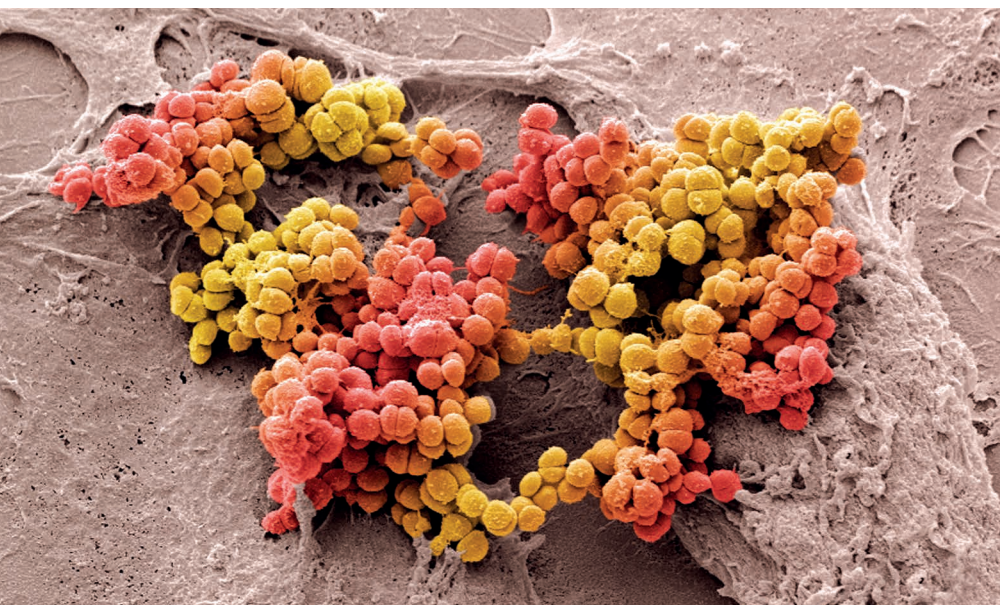
Gallo's team has since discovered at least 14 strains of *S. epidermidis* that produce antimicrobial compounds, which kill microbes that trigger inflammation. Some of those compounds inhibit the growth of *S. aureus*, a species of *Staphylococcus* that can cause serious infections. Certain strains of *S. aureus*, including methicillin-resistant *S. aureus*, have become resistant to multiple antibiotics.

The discovery that species of *Staphylococcus* could offer protection from others holds particular promise for treating, or even preventing, atopic dermatitis, the most common form of eczema. The condition, which affects more than 18 million people in the United States and up to 30% of children in industrialized countries, causes itchy, red patches on the skin that have an impact on quality of life and increase the risk of infection. Eczema is also thought to have a bacterial component: when patches develop, the number of *S. aureus* bacteria on affected regions increases, and there is a decline in the overall diversity of the skin microbiota. People with more severe cases of eczema experience larger increases in the number of *S. aureus* bacteria on their skin.

That rise might, in part, be responsible for exacerbating eczema flare-ups. In 2017, Segre, Heidi Kong at the US National Cancer Institute in Bethesda, Maryland, and their colleagues isolated *S. aureus* from the skin of 18 children with or without eczema⁵. The team then allowed bacteria from both groups to colonize healthy mice. Mice that received *S. aureus* from children with eczema went on to develop eczema-like inflammation and thickening of the skin.

If *S. aureus* does cause flare-ups, then fighting that microbe might help to treat eczema. Gallo has developed a cream that incorporates strains of *Staphylococcus* found on human skin that inhibit *S. aureus* by producing an antimicrobial compound. In a randomized, double-blind trial of the cream in 11 people with eczema, all of whom were deficient in the inhibiting bacteria before receiving Gallo's intervention, a single application led to a more than 90% reduction in the number of *S. aureus*

"It speaks to a window of opportunity for establishing a healthy relationship with skin bacteria."



The bacterium *Staphylococcus aureus* (red and yellow) is found on human skin.

STEVE GSCHWEISSNER/SPL

on the skin of all participants. “It worked fantastically,” Gallo says. “It was basically the first time in humans that a microbial transplant on the skin was therapeutically useful.”

Gallo says his latest results show that applying the cream twice a day, every day for a week reduces *S. aureus* numbers by more than 99%, with symptoms declining in severity by 20–30%. MatriSys Bioscience in La Jolla, California, a company co-founded by Gallo, is aiming to bring the cream to market in two to three years.

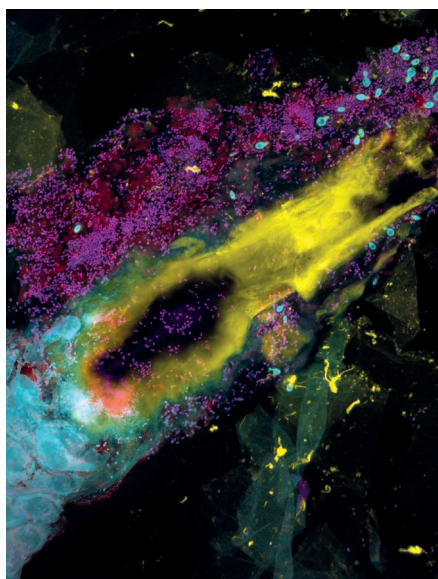
Such treatments could help to tackle more than just skin conditions. People with eczema often develop asthma and allergies — previously thought to be a sign of overall immune-system dysfunction. Some researchers now suspect that microbes living on shed skin cells — a major component of dust in the home — exacerbate these allergic reactions. If this hypothesis is correct, it might be possible to alter the skin microbiota in a way that could ease eczema while also making flaked-off skin less likely to trigger the immune system when inhaled. Or, Segre adds, it might be possible to predict which treatment will work best or when someone might be about to experience an eczema flare-up.

There are hints that fungi might also help to protect skin from eczema, says Thomas Dawson, a pharmacologist and chief executive of skin-product consultancy Beauty Care Strategies in Singapore, who previously spent more than 15 years using skin-microbiome research to develop anti-dandruff products. One clue lies in the point at which certain fungi colonize skin: *Malassezia*, comprising 17 species, peaks in babies, who tend to have greasy skin, and again at puberty — a time when skin becomes oilier and eczema becomes less common. Another clue comes from a 2016 study by researchers in Singapore, China and the United States, who found that the *Malassezia* population decreases as *S. aureus* numbers rise during eczema flare-ups⁶. In 2018, Dawson and his colleagues found a potential mechanism for this relationship, in which some types of the fungus secrete enzymes that digest *S. aureus* biofilms⁷.

The role of *Malassezia* in skin health is far from certain; however, there is evidence to suggest that fungi also cause certain other skin conditions. Clarifying the details could lead to the development of strain-specific anti-fungal treatments, or probiotics (live microbes) that when applied to skin help to engineer a health-promoting balance of fungi. “By no means is the hypothesis proven” that skin fungi can be protective, says Dawson, who is also president of the Skin Research Society Singapore. “But the picture is starting to become more clear.”

ACNE ACME

Work on eczema is at the forefront of research on the skin microbiota, but other skin conditions such as acne could also benefit from a growing understanding of skin microbes. For decades, scientists have known that the bacterium *Cutibacterium acnes* (formerly *Propionibacterium acnes*) thrives on the skin



Hair follicles host bacteria (pink) and yeast (teal).

of people with acne, which affects up to 85% of teenagers. Antibiotics that target *C. acnes* have therefore long been an effective treatment for the skin disorder.

But initial attempts to solidify a link between acne and the skin microbiota were disappointing, says Huiying Li, a bioinformaticist at the University of California, Los Angeles. When she used next-generation sequencing to compare the microbial make-up of healthy and acne-prone skin, both tended to have similar relative abundances of *C. acnes* bacteria, which suggested that the microbiota might not play a part in acne after all.

However, Li found population-level differences in strain composition that distinguish skin with acne from healthy skin⁸. So far, her team has sequenced 70 of more than 120 known strains of *C. acnes*, revealing genes that are key to the bacterium's virulence, as well as potential explanations for why some strains are associated with acne. Such strains, she suggests, produce much greater amounts of porphyrins, which are molecules that can trigger inflammation in skin cells. Li predicts the development of medications that could help people with acne by eliminating virulent strains of *C. acnes* or by targeting porphyrins. Some companies are already selling products that claim to tackle acne by killing *C. acnes* and restoring balance to the skin microbiota, even though evidence to support the approach is limited.

Researchers also hope to use skin microbiota therapies to target a potentially life-threatening disease: skin cancer. Earlier this year, Gallo and his colleagues reported that certain strains of *S. epidermidis* produce a molecule called 6-*N*-hydroxyaminopurine (6-HAP) that has anti-tumour properties in mice⁹. Injections of 6-HAP slowed the growth of aggressive melanomas in the animals. And the application of *S. epidermidis* that produce 6-HAP to mouse skin reduced the number of tumours that formed in response to ultraviolet

radiation. The researchers also found that some people carry 6-HAP-producing strains of *S. epidermidis*. These people might have some natural protection against skin cancer, raising the possibility that doctors could identify people without such bacteria, who might be at higher risk, to take preventive measures.

Researchers speculate that, in the next five years, altering the skin microbiota might become part of routine interventions such as those that aim to protect people from skin infections, especially in hospitals, where drug-resistant *S. aureus* is a growing problem. To understand more about how *S. aureus* infections are acquired, researchers injected mice and zebrafish with *S. aureus* and common, non-pathogenic types of skin bacteria¹⁰. Those bacteria made the animals more susceptible to *S. aureus*, reducing the number of bacteria that were required to cause an infection. According to Oh, this means that a person's probability of picking up a *Staphylococcus* infection might depend on the composition of their skin microbiota. “That could explain why some individuals are more susceptible,” Oh says. “They have different commensal microbes surrounding the context of the pathogens.”

In her lab, Oh has been systematically putting together combinations of bacteria that live on skin, to see how various species fare when trying to colonize particular microbiota. It is one of many basic questions that scientists need to address before they can safely and effectively alter the skin microbiota to improve health.

Looking beyond bacteria to fungi and other microbes will also be important. Segre and her colleagues have already published a study on the often underestimated amount of viruses that are found on skin and that could affect a person's vulnerability to warts and other skin conditions¹¹.

As hype builds, researchers are casting a wide net in their studies of skin microbes. As a follow-up to his hospital microbiota project, Gilbert has been working with NASA to study communities of microbes on the International Space Station. His findings suggest that bacteria there act much like those in hospitals on Earth, moving from people to their environment, and back. Such insights — combined with other, ongoing research on the microscopic life that lives upon us — could help to transform health care, on Earth and beyond. ■

Emily Sohn is a freelance journalist in Minneapolis, Minnesota.

1. Lax, S. *et al. Sci. Transl. Med.* **9**, eaah6500 (2017).
2. Oh, J. *et al. Cell* **165**, 854–866 (2016).
3. Scharschmidt, T. C. *et al. Immunity* **43**, 1011–1021 (2015).
4. Lai, Y. *et al. Nature Med.* **15**, 1377–1382 (2009).
5. Byrd, A. L. *et al. Sci. Transl. Med.* **9**, eaal4651 (2017).
6. Chng, K. R. *et al. Nature Microbiol.* **1**, 16106 (2016).
7. Li, H. *et al. J. Invest. Dermatol.* **138**, 1137–1145 (2018).
8. Fitz-Gibbon, S. *et al. J. Invest. Dermatol.* **133**, 2152–2160 (2013).
9. Nakatsuji, T. *et al. Sci. Adv.* **4**, eaao4502 (2018).
10. Boldock, E. *et al. Nature Microbiol.* **3**, 881–890 (2018).
11. Tirosh, O. *et al. Nature Med.* <https://doi.org/10.1038/s41591-018-0211-7> (2018).



To scientists such as Vissers, it's clear that the skin needs to be fed with nutrients such as vitamin C from within. Although our skin is exposed to the outside world, it's relatively inaccessible for external nutrients, says John Casey, who was vice-president for bioscience research at Unilever in London for ten years. Pollutants in the environment can make their way through (see page S89), but, says Casey, who is now retired, "nutrients important to fuel and feed the skin are entirely different". Essential compounds, such as vitamins, sugars, peptides and minerals, are often large and water soluble. "Things that you apply from a topical will not pass that barrier. They will not get down to the living layers of the skin," he says.

A growing body of research, on everything from anti-ageing strategies to cancer risk, suggests that diet might be key to skin health. However, the practical details are unclear. The best diet advice for ensuring healthy skin aligns with general guidelines: eat a varied diet full of fruits, vegetables and other unprocessed food. Now, researchers must translate their findings into specific advice about which nutrients, in which quantities and combinations, will ensure skin health. So far it is proving a difficult proposition.

ALPHABET SOUP

Vissers has been studying the role of vitamin C in immune function, mood, mental health and even cancer for more than a decade. Now, she's beginning to investigate links between vitamin C consumption and the levels found in the bloodstream and the skin. "The skin goes to great lengths to take up vitamin C," Vissers says. She compares it to a vital link in a long chain. "It influences so many processes that without it, many things are going to falter." Vitamin C is necessary for protection against sun damage in the epidermis, where it mops up free radicals produced by UV rays. It may also be involved in the maturation of keratinocytes, the cells that make up the epidermis.

In the thick inner dermis, vitamin C is needed to produce and maintain collagen, the spongy protein that gives skin its underlying structure and plump appearance. It also increases proliferation and migration of fibroblasts, the cells responsible for collagen production, and regulates signalling pathways related to inflammation, aiding wound healing.

People with diets that lack vitamin C can be at risk of scurvy, a condition that can result in exceedingly dry, brown-tinged skin, excessive bruising and slow-healing wounds (S86). But until now, scientists had little information about the link between dietary and skin vitamin C in healthy individuals. Vissers and her team have unpublished data showing that the amount of vitamin C a person eats maps directly onto the vitamin C content in their skin. Therefore, "you can boost vitamin C in compartments of the skin by improving your diet," Vissers says.

Vissers isn't alone in probing the links between nutrition and the skin. Many studies have focused on the goal of keeping skin

NUTRITION

Edible skin care

Eating well could be the best defence for our skin. But which vitamins and nutrients will yield the healthiest glow?

BY SARAH DEWEERDT

Biochemist Margreet Vissers shares a common enemy with skincare companies: the highly unstable free radicals that damage cells and attack DNA. Her latest work, however, which looks at the effects of vitamin C on skin health, is not focused on developing new creams or lotions. "I've been known to say to cosmetic companies, 'you know you'd probably be better off eating your product than rubbing it on,'" says Vissers, who

heads the Centre for Free Radical Research at the University of Otago in Christchurch, New Zealand.

Such a comment reflects a growing awareness of the role of nutrition in skin health. Skin is the largest organ in the body, comprising about 10–15% of body weight. It helps protect the body from dangers such as ultraviolet rays, pollution and infections, and it constantly renews itself — the outermost layer, the epidermis, remakes itself every month. All of that requires a constant flow of energy and nutrients.

looking youthful — plump, dewy, firm and unwrinkled. Scattered studies of cells in the laboratory, animal models and a few human trials also support roles for a variety of nutrients in preventing skin ageing. These include vitamins, not just C, but also vitamin D and E; carotenoids, such as β -carotene, lutein and lycopene; and plant-based chemicals found in foods that range from soya and turmeric to chocolate and green tea.

But despite researchers' mechanistic knowledge of how compounds such as vitamins and minerals might work, scientists still don't know much about the optimal intake to stave off skin ageing. One observational study¹, which included more than 4,000 women in the United States aged 40–74, suggested that a diet rich in vitamin C and linoleic acid (an omega-6 fatty acid found in nuts, seeds and vegetable oils) is associated with younger-looking skin. Another study², this one of 716 women in Japan, suggested green and yellow vegetables might be the best choice.

However, such studies are inconsistent: in the US study, women who consumed lower levels of fats had younger looking skin, whereas in the Japanese study this was true for those who ate more.

The result is a cacophony of claims that can be difficult for consumers to sort through.

One of the most rigorous evaluations of nutritional supplementation to fight ageing came in 2014, when Casey and his colleagues at Unilever developed a nutritional supplement and tested it in a randomized controlled trial³. The supplement combined five ingredients, each of which had promising anti-ageing properties.

Their Strength Within Anti-Wrinkle Supplement included antioxidants (vitamins C and E), as well as lycopene, which absorbs UV light and soaks up free radicals. It also contained soya isoflavones that Casey says boost collagen production, at least in culture. The final ingredient was a fish-oil supplement, rich in omega-3 fatty acids that upregulate collagen synthesis and have anti-inflammatory properties.

At the end of a 14-week study in 159 women, those who took the supplement daily had reduced wrinkle depth and skin that contained more freshly synthesized collagen compared with the control group. With these data in hand, a Unilever subsidiary called Dove Spa launched the supplement in 2011. But there was little marketing effort, Casey says, and two years, later the pills were removed from the market when the subsidiary was sold. It has since been relaunched by Ioma, a cosmetics company in Paris, as Collagen Renew.

SUN SIGNS

Emerging evidence suggests that nutrition may help to prevent melanoma. Multiple studies point to vitamin D as a potential defence against this aggressive skin cancer, which results from exposure to UV light.

In vitro studies have shown that vitamin D dampens proliferation in melanoma cell lines⁴.



Tissue samples are processed for vitamin C analysis at the University of Otago in New Zealand.

And epidemiological studies have found that people with more advanced melanomas tend to have lower levels of vitamin D in their blood than those with less advanced tumours⁵.

Eggs, meats, mushrooms and fortified dairy products all contain vitamin D. But when bathed in sunlight, skin can make the vitamin itself. Researchers have long known that a little bit of sun exposure is healthy for the body for a variety of reasons, although too much can prove harmful. But now they're find-

“Sunlight and vitamin D could be really important for melanoma outcomes.”

ing that moderate sun exposure might protect against the very harm that excessive exposure causes. “Sunlight and vitamin D could be really important for melanoma outcomes,” says Michael Kimlin, a cancer-prevention researcher at the University of the Sunshine Coast in Brisbane, Australia.

Kimlin and his team showed⁵ that people with melanoma and low vitamin D levels were more likely to have thicker tumours, which generally have a worse prognosis. By measuring vitamin D levels at diagnosis, the team was able to exclude the possibility that low vitamin D levels were due to people with more severe melanomas being more diligent at staying out of the sun in the wake of their diagnosis.

But it's still unclear whether the vitamin itself is the protective factor. Levels of vitamin D in the blood could be a marker for another protective effect of sunlight, or some other sunlight-influenced nutrient altogether.

For people with average skin-cancer risk, these findings don't change the commonsense advice to wear sunscreen and get outside. People rarely apply enough sunscreen for it interfere with the body's ability to make vitamin D. “Time and time again our studies in Australia show the people who sun-protect the most actually have

the highest levels of vitamin D,” because they also tend to be more active and spend more time outside, he says.

But for those with a high melanoma risk, or those who have already been diagnosed, this line of research suggests that oral vitamin D supplementation could be a good strategy. A randomized trial of vitamin D supplementation in high-risk individuals might be worthwhile, Kimlin says. Several large randomized trials are already underway to investigate whether this strategy could help prevent other forms of cancer. But preliminary results suggest that even though vitamin D levels have been linked to cancer protection in epidemiological studies, supplements might have little effect. “When you start taking nutrients out on their own, and you start to look at anticancer properties, it doesn't necessarily replicate what we see in the observational studies,” Kimlin says.

Vitamin D supplements are so ubiquitous that it's difficult for researchers to gauge their anti-cancer effect. And differences in individual biology can obscure patterns. Kimlin's ongoing research aims to determine how differences in the vitamin D receptor gene affect melanoma risk. But whether researchers are talking about wrinkles or melanoma, the sticking point is the same: the leap from general healthy lifestyle advice to specific recommendations about a particular nutrient remains a challenge, and not one that will be solved anytime soon. ■

Sarah DeWeerd is a freelance science journalist based in Seattle, Washington.

1. Cosgrove, M. C., Franco, O. H., Granger, S. P., Murray, P. G. & Mayes, A. E. *Am. J. Clin. Nutr.* **86**, 1225–1231 (2007).
2. Nagata, C. *et al. Br. J. Nutr.* **103**, 1493–1498 (2010).
3. Jenkins, G., Wainwright, L. J., Holland, R., Barrett, K. E. & Casey, J. *Int. J. Cosmetic Sci.* **36**, 22–31 (2014).
4. Slominski, A. T. *et al. J. Steroid Biochem. Mol. Biol.* **177**, 159–170 (2018).
5. Wyatt, C., Lucas, R. M., Hurst, C. & Kimlin, M. G. *PLoS One* **10**, e0126394 (2015).



BEYOND THE BIOLOGICAL

Skin-like electronics that stretch and sense will create a way to monitor vital signals in the long term, and build prosthetics with a sense of touch.

BY KATHERINE BOURZAC

Sitting at her desk at Stanford University in California, chemical engineer Zhenan Bao holds a stretchy sheet of electronics. It is clear, squishy and less than a millimetre thick. When she holds it up to the light, the metallic elements that make up the electronic interconnections become visible. This sheet can do much the same thing as a rigid circuit board, but it wrinkles and gives like skin.

“We view the skin as a wearable electronic system,” Bao says. “It has all the components you want to mimic to make better prostheses, wearable sensors and smarter robots.” Unlike conventional electronics, which are rigid and brittle, skin is resilient. It’s stretchy and self-healing. And skin contains a remarkable, integrated network of energy-efficient sensors that pick up pressure, temperature and more. It seems almost a natural leap, then, for researchers such as Bao to suggest using skin-like electronics to give prosthetics a sense of touch.

A slew of chemistry and materials-science advances over the past 15 years have enabled researchers to come close to mimicking many of skin’s properties — its ability to self-heal, its stretchiness and its sensing capabilities. Now, scientists need to find ways to pull those advances together in a single design. And they must show that such skin-like devices can do more than single time-point medical measurements, as

well as making the artificial skin itself robust enough to wear over long periods of time. And it is not just human skin providing inspiration, researchers have been trying to mimic how creatures such as octopuses change their appearance (see ‘Underwater inspiration’).

NEW SKINS

Human skin is a sensitive, sophisticated and robust organ. It is water-resistant and heals when cut. Its multitude of mechanoreceptors detect sensations such as vibration, pressure and texture — they’re sensitive enough to detect the faint pressure of a breeze or alighting flies. The tight coupling of skin sensors with the peripheral nervous system is responsible for our reflexes and allows us to pick up objects of different weights, shapes and textures without conscious thought. Such properties might seem unremarkable, but for a person with an inert prosthetic hand or an electrical engineer trying to make resilient, low-power devices, human skin is a wonder.

Early on, researchers working on skin-like electronics focused on robotics applications, says Takao Someya, an electrical engineer at the University of Tokyo. Robots with a sense of touch could perform more-complex tasks and are less likely to break things or hurt someone.

But as Someya and his contemporaries

Flexible electronics are enabling prostheses that can feel.

worked to give their robots some resemblance of skin, they hit a wall. Flexible electronics have more give than rigid ones, but they still restrict the range of motion when wrapped around finger- or elbow-like robotic joints. “We quickly realized how important it was to introduce mechanical stretch. Without stretchability it’s impossible to apply the electronic skin to moving parts like joints or to curved surfaces,” says Someya, who made the first flexible, large-area pressure sensor in 2003. As those electronics became more stretchable, researchers realized that the materials could be made biocompatible and applied to the skin itself.

In 2011, materials scientist John Rogers, now at Northwestern University in Evanston, Illinois, made what he called epidermal electronics: thin sheets of circuits, the mechanical properties of which were engineered to match those of human skin. Using a suite of mechanical- and materials-engineering techniques, Rogers made rigid silicon — the electronic industry’s material of choice — compatible with flexible and stretchable surfaces.

Rogers still uses the same basic set of techniques. Researchers in his lab etch thin silicon components and then use a specially designed stamp to pick them up and transfer them onto rubber-like material. The rigid components sit on ‘islands’ that have been mechanically engineered to protect the silicon from mechanical strain. The silicon electronics — including light-emitting diodes, electrodes and sensors — are connected by spring-like metal wires made using kirigami, a form of origami that uses both cutting and folding.

Rogers is now focused on medical applications, including prosthetic limbs. To do this, he’s collaborating with Levi Hargrove, director of the Center for Bionic Medicine at the Shirley Ryan AbilityLab, a rehabilitation research centre and hospital in Chicago, Illinois. The technology has made great strides in the past decade, Hargrove says. Prosthetic arms and hands now exist that have articulated fingers, rotating wrists and elbows. They can lift heavy weights, but they lack the sensors needed to detect, for example, that a cup of coffee is scalding hot, and the nerves to rapidly transmit the corresponding reaction signal back to the hand: put it down!

Hargrove and Rogers are collaborating on a system that they hope will enable someone to control a prosthetic hand more like they would a real one. As the wearer imagines moving their wrist or closing and opening a fist, their muscles contract. Some prosthetics (mostly in research labs) use electromyography — electrodes placed on the muscles at the site of the amputation — to pick up on faint electronic traces of contractions. An on-board processor then interprets these electronic signals to control the prosthetic. The approach is promising but awkward, requiring constrictive rubber cuffs to hold 2-millimetre-high electrodes in place over the muscle. The electrodes can irritate the skin and the cuffs can feel uncomfortable and cause people to sweat, degrading the quality of the electrical signal. The accuracy and detail of the electrical readings also need improving.

Together, Hargrove and Rogers are aiming to improve both the comfort and accuracy of these systems. Their skin-like electromyography patch is just tens of micrometres thick, lightweight, has holes for sweat to evaporate and doesn’t move around. The researchers have tested it on three people so far. Each epidermal electronic patch is just a few square centimetres and, when placed on the residual limb, allows users to control their robotic, prosthetic hands.

The main challenge that they now have to overcome is the skin itself. “You’re limited by exfoliation,” Rogers says. “The build-up of dead skin cells disrupts adhesion and the facility of electronic measurements.” As a result, the patches themselves only last a week or two, at most. “We’d like to do months,” he says. Every time a patch is replaced, users must readjust: small changes in location means that it picks up signals from a slightly different set of muscles. And each time, the user has to relearn how to control their robotic hand. Rogers would like to collaborate with biologists to solve the problem, and hopes that they could help him to form an enzyme-saturated adhesive to digest the dead cells.

These stretchy patches can pick up not just electrical signals but chemical and physical ones, too. Some researchers, including Rogers, are using them to measure metabolites such as lactic acid — a sign of muscle fatigue — or glucose in sweat. Someya is working with dermatologists at



A flexible circuit designed to capture electrical signals from muscle activity.

the Keio University School of Medicine in Tokyo to develop electronic-skin patches that could assess allergies by measuring rates of evaporation from the skin. Dryness is often a symptom of allergic symptoms. And he’s developing ultra-lightweight sensors that look like a temporary tattoo to monitor blood oxygen and heart rate, along with wearable displays to visualize the data.

“WE VIEW THE SKIN AS A WEARABLE ELECTRONIC SYSTEM.”

Someya says one of the most promising potential applications of electronic skin is to monitor vital signs and chronic medical conditions over years or months. In tests, the patches take measurements that are at least as accurate as conventional medical equipment, but they are much more comfortable — people can’t even feel them. However, long-term monitoring won’t be possible until researchers can improve the technology’s usability. One approach is to make the sensors disposable, like a sticking plaster. But each time the sensor is reapplied, the data computation has to be recalibrated. So far, most studies have been short, but researchers need long-term data to develop their algorithms — a problem they hope will soon be addressed.

FINDING THE NERVE

Sensors that run continuously, such as those of electronic skin, pose a huge information-processing challenge — whether they’re worn for months or just a day. To handle information as conventional computers do, they would have to send each data point to a central processing unit, an inefficient and battery-draining prospect.

To get ahead of this potential data onslaught, Bao is finding inspiration in the way mammals and other animals process information, which is low power and error tolerant. Bao has described² what she calls an ‘artificial nerve’, one that mimics skin mechanoreceptors, neurons and the synapses that connect them to the spinal cord. This artificial nerve is made from intrinsically stretchy, electrically conductive polymers. The circuits within are also bioinspired. The artificial nerve converts readings from pressure sensors into electrical signals similar to those sent by real nerves. Then, instead of processing every signal individually, these signals are added up by a transistor that mimics the biological synapse. This has two benefits. It could help to ensure biocompatibility. And it results in a system so exquisitely sensitive that it can detect and read something as small as the tiny bumps of braille.

Closing the loop between the skin and the nervous system — so that a

CEPHALOPODS

Underwater inspiration

Mammalian skin isn't the only muse for materials scientists. Many are enthralled by cephalopods — the group that includes octopus (**pictured**), squid and cuttlefish — which can change their appearance in microseconds to match a colour or texture in the environment, or create a disorienting pattern to confuse predators. Octopus-like, colour-changing materials could be used in military camouflage, electronic displays, or even in art and cosmetics. Biologists are working to understand how cephalopod skin works and material scientists are trying to mimic it.

The cephalopods' skin has multiple layers. A white bottom layer provides a reflective background that acts as a blank canvas. Above that is a layer of reflective iridophore cells; some of these can be switched on and off by muscle contractions that hide or reveal their sparkly contents. Next are chromatophores, which expand or contract to reveal patches of coloured pigments. And on the surface layer, octopus and cuttlefish have papillae that pop up to create a rough texture, or go down for a smoother appearance.

But despite what they know about how cephalopods control their appearance, scientists are just beginning to understand how the creatures monitor their environment so astutely. The first clues came in 2015, when researchers found⁴ light-sensing opsin proteins in cephalopod chromatophores, indicating that the animals can 'see' with their skin.

So far, researchers have only been able to mimic the most basic parts of the cephalopod's adaptive camouflage. In 2014, scientists layered a sheet of light sensors and one of colour-changing elements, with pixels that alternate between black and white, that could mimic the underlying image⁵. Such technology could ultimately be used to make adaptive camouflage coatings for vehicles or even clothing. Other researchers are using a similar design — a material that senses the background and matches it — to make cephalopod-inspired thermal camouflage. **K.B.**

wearer could both 'feel' with a prosthetic limb and control it as naturally as they could a real one — will be challenging, says Dae-Hyeong Kim, a chemical engineer at Seoul National University. Sensing is just one part of such a system. Nerve signals go both ways, to the brain and back again, and when someone can feel what the prosthetic senses, they can also react — both voluntarily and involuntarily.

Bao is particularly interested in mimicking natural reflexes. Without thinking about it, for instance, we quickly withdraw our hands from something painful. That's thanks to sophisticated processing of sensory data: first, so-called afferent nerves transmit information from sensors in our fingertips to the spinal cord. Then, efferent nerves rapidly carry muscle-control signals back to the hand. Bao's artificial nerves already do something similar, adding up mechanical signals and converting them into an actuating one. Bao's team even made a crude demonstration of the possibilities by connecting the artificial nerve, which was programmed with a reflex designed to control muscles, to a cockroach's leg. When the artificial nerve is pressed, the leg twitches as if it has been touched.

"WE'VE DEVELOPED A MATERIAL WITH TISSUE-LIKE SOFTNESS."

This work is fairly preliminary, and Bao will need to show that it works in people. She's hopeful that a future version of this squishy artificial nerve will be biocompatible and could be connected to nerves in a person's residual limb. "We've developed this material with tissue-like softness," she says, a suppleness designed to ensure comfort. But, Kim notes, researchers still have to work out how to connect "the abiotic part with the biotic part" — the synthetic with the biological — over the long term.

Beyond ensuring biocompatibility, designers must also ensure that these systems have a reliable power source, says Xue Feng, director of the Center for Flexible Electronics Technology at Tsinghua University in Beijing. Currently, electronic skins must either be tethered to a power source through a wire — requiring the person or robot to also be tethered — or packaged with a battery.

Feng thinks that various energy-harvesting techniques could help. Removing the need for a wire or a battery would mean that artificial skin is feasible for those who have a prosthetic. Feng's group has experimented with using materials that convert mechanical energy into electrical voltage. This could mean a bend of an elbow or the strike of a footfall is converted into a jolt of power. And Someya's group has been working on stretchable, skin-like solar cells that, when incorporated into the top layer of an electronic skin, could supply power during the day.

Meanwhile, Bao is already working on more bioinspired designs, including electronic materials that have the ability to heal. In August, her group described³ a self-healing electronic system: a wearable heart-rate monitor and simple display. The stretchy system relies on the movement of polymers to 'heal' wounds. The conductive polymers flow back together after they're cut, and could mean that if a future robotic prosthetic gets a cut, it can repair itself.

"We read how the skin works, and then we try to see how we can mimic that," Bao says. If researchers can make electronics not only look like skin, but also function like it, this could make a big difference for people who wear prosthetics. "At the moment most prosthetic hands are just for the look," she says. "People wearing them are enthusiastic about being able to manipulate objects more precisely, or even regain their sense of touch." ■

Katherine Bourzac is a science journalist based in San Francisco, California.

1. Kim, D.-H. *et al. Science* **333**, 838 (2011).
2. Kim, Y. *et al. Science* **360**, 998–1003 (2018).
3. Son, D. *et al. Nature Nanotechnol.* **13**, 1057–1065 (2018).
4. Kingston, A. C. N. *et al. J. Exp. Biol.* **218**, 1596–1602 (2015).
5. Yu, C. *et al. Proc. Natl Acad. Sci. USA* **111**, 12998–13003 (2014).

PERSPECTIVE



The eye of the beholder

Clinicians and researchers must learn to talk about and treat vitiligo without alienating a growing chorus of patient advocates, says **John Harris**.

As a clinician–scientist who studies and treats the skin disease vitiligo, I am committed to improving the lives of my patients. The condition, which is characterized by white patches on the skin, results from the immune-mediated destruction of melanocytes — cells that produce melanin, the pigment that gives skin its colour. Vitiligo affects about one in every hundred people worldwide, without bias towards gender, race or geographic location. But it could be even more prevalent: many patients tell me that they hadn't sought clinical care previously (and therefore weren't included in official data) because they were unaware that treatment options exist.

At the clinic, I care for people who do seek treatment. In the laboratory, I work to better understand the mechanisms by which vitiligo arises and then progresses to develop potential treatments. And in the vitiligo community, I act as an advocate for patients to ensure that their treatment costs are covered, and to help reduce the stigma that is associated with the disease. Although these activities might seem synergistic, they have collided in unexpected ways. While my patients and millions of others around the world are clamouring for a cure, many so desperate that they go to great lengths to hide their symptoms, some people with the condition are demanding that it be accepted as part of everyday diversity. The clash could have a lasting effect on vitiligo research, funding and treatment.

The impact of vitiligo on a person's quality of life is comparable to that of skin conditions such as psoriasis and atopic dermatitis (eczema). Like those conditions, vitiligo is something that we should take seriously and seek to treat. But the condition is often overlooked by both clinicians and researchers, who dismiss it as being cosmetic. Vitiligo's impact and the care it demands is overwhelming, yet I battle daily with insurance companies to get treatment coverage for my patients, and with funding agencies to secure research support.

Depending on where they live, people with vitiligo can be admired for their stunning beauty or rejected for their shocking disfigurement. Many cultures attach a strong social stigma to the disease. One of my patients told me that on the flight he took to my clinic, the woman next to him asked to change seats because of the spots on his arms. In south Asia, and in India, in particular, vitiligo was once confused with the infectious disease leprosy, which is endemic in the region. Even today, India's culture of arranged marriage still ostracizes those who have vitiligo; not only those with the disease but also their siblings can be eliminated as marriage prospects. And in the United Kingdom, a man who was originally from Pakistan reportedly asked for advice on how to arrange amputation of part of his arm, which showed signs of vitiligo. He said that his family would accept him with just one arm, but not with the condition.

There is, however, an alternative narrative: one that hails the disease as something to be celebrated. Art exhibitions, advertisements and television shows feature models and actors with vitiligo, serving as an acclamation of the condition and the patterns that it imprints onto the skin of those affected. Some people even resist calling vitiligo a disease, possibly because the term subverts acceptance. Although this

body-positive standpoint seems to be that of a vocal minority, it receives the majority of media attention and helps to raise awareness of vitiligo worldwide. But it also creates a conflict, in which those who want to receive treatment feel undermined by those who do not, and vice versa. Similar dual narratives — in which seeking treatment or cosmetic solutions for a condition puts people at odds with others who accept it openly — exist in the albinism, dwarfism and deaf communities.

Such conflict is uncommon in other immune-mediated skin conditions. Some, including psoriasis and eczema, can be accompanied by discomfort in the form of itching or pain. But others, such as alopecia areata, typically are not, so this alternative perspective does not stem solely from lack of discomfort. And although only a limited number of treatments are available for vitiligo, many other diseases have even fewer effective therapies. The greater visibility of models and other individuals with vitiligo might have helped many in the community to embrace the

disease. But when singer Michael Jackson struggled publicly with vitiligo more than 20 years ago, he received a very different reception. Perhaps a combination of all these things, as well as cultural evolution, has created room for this perspective.

But no matter the reason, this second narrative has implications for the first. If vitiligo is beautiful, perhaps research on its treatment is unnecessary. Some might feel that people who seek care or feel devastated by their diagnoses imply that others, even those who accept it, should be ashamed of their appearance. Increased resistance to treatment and forced acceptance of the disease could impede the research effort.

Both perspectives are valid, healthy and appropriate. And although they seem to contradict each other, they confront the reality that

vitiligo cannot be ignored. As a clinician–scientist who has committed a considerable part of my career to treating and studying vitiligo, I straddle the line: I support patients who love their appearance and decline treatment and comfort those whose spots move them to tears.

The field is making progress. Clinicians are getting better at caring for people with vitiligo. Basic and translational research have provided insight into how vitiligo arises and real options for improved treatments, which are now being tested in clinical trials. Pharmaceutical companies have noticed that vitiligo represents an enormous, unmet clinical need.

Clinicians, researchers, patient advocates, people with vitiligo and their caregivers must set aside personal biases and seek to understand the effects that vitiligo has on all those it touches. Better yet, they should join the conversation. By working together, each of these groups can help people with vitiligo who feel alone and powerless, and empower those who call attention to the condition's unique beauty. There is now real hope for people with vitiligo, in terms of public acceptance of the condition and new, advanced treatments. And that is what's most important, no matter how you look at it. ■

John Harris is director of the Vitiligo Clinic and Research Center at the University of Massachusetts Medical School in Worcester.
e-mail: john.harris@umassmed.edu

PEOPLE WITH
VITILIGO CAN BE
ADMIRED
FOR THEIR STUNNING
BEAUTY OR
REJECTED
FOR THEIR SHOCKING
DISFIGUREMENT.

Author: Almirall R&D

Research partnerships in skin health



The innovation process in pharmaceutical research is continuously evolving. In the past, companies used to work in a closed model where new ideas and knowledge came mainly within house. This model is no longer sustainable and the pharmaceutical industry is increasingly opening up the innovation model. Nowadays, collaborating with all stakeholders involved in the research process, such as academia, biotechnology firms, other pharmaceutical companies and funding bodies, is essential for a successful business and to transform innovative ideas into valuable solutions for patients.^{1,2}

At Almirall, a global pharmaceutical company focused in medical dermatology, we firmly believe in partnerships to share knowledge, science and efforts, and to build synergies that accelerate the R&D process. Therefore, as part of our strategy to find treatments that are tailored and differentiated from the current standard of care, we are implementing external innovation initiatives to identify therapies, technologies and devices that address unmet medical needs in dermatology. Different types of collaborations implemented at Almirall include public-private partnerships, private-private partnerships, funded projects, and crowdsourcing initiatives (Fig. 1). Recent examples of research partnerships in dermatology are illustrated below.

Research partnerships

In 2017, Almirall and LEO Pharma joined forces to establish a public-private partnership to advance the understanding and treatment of skin diseases by setting a new standard for skin sampling. The aim of the project is to develop and clinically validate a painless, minimally invasive skin sampling method that enables accurate and comprehensive biomarker analysis in clinical trials and exploratory research.

The research is being conducted at the Hospital Clinic of Barcelona (Spain), the Technical University of Denmark and the University of Bath (UK). The methodology will be validated in selected patient populations. Final research results will be published in peer reviewed journals.

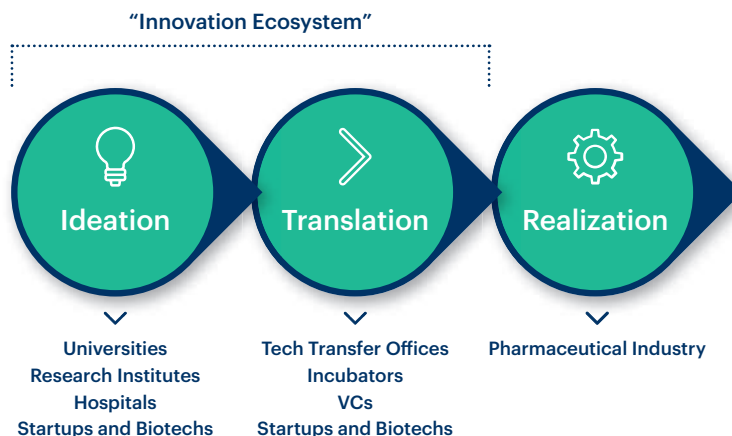


Figure 1: Different Sources of Innovation at Early Stages

The research collaboration between Almirall and Mercachem focused on the development of oral cytokine blockers for the treatment of inflammatory skin diseases. Within this project, which started in 2017, Mercachem is performing iterative optimization campaigns towards the identification of oral cytokine blockers for Almirall to further develop them. The cytokine blockers programme has its origin in the collaboration of Mercachem with Viperger using its unique DNA-encoded library and Binder Trap Enrichment® (BTE) assay technology (www.viperger.com). This breakthrough technology allows the identification of small molecules that can block protein-protein interactions. It opens up a whole new category of drug targets of great interest in dermatology.

Very recently, Almirall and Evotec have announced a research collaboration to discover and develop first-in-class therapeutics through a novel approach to disrupt cell signalling that is expected to deliver highly potent and durable treatments for debilitating dermatology diseases such as psoriasis and atopic dermatitis.

Reaching the crowd: AlmirallShare

A key factor to successfully access external innovation is to reach the wider scientific crowd. With this in mind, Almirall launched AlmirallShare (sharedinnovation.almirall.com) a web open innovation platform created to facilitate collaborative projects with scientists world-wide and find innovative solutions to address unmet needs in skin health.

AlmirallShare addresses different dermatological challenges in a flexible and customized manner. It offers several collaboration opportunities including the development of translational research models for understanding the pathology of skin diseases, the identification of novel targets for dermatological disorders, and the identification of new uses in dermatology for existing advanced compounds. Additionally, it may offer funding opportunities and mentorship for collaborators.

Scientists from universities, research centers, hospitals, start-ups, biotechnology firms, and pharmaceutical companies all over the world with an interest in dermatological research are invited to participate. All submitted proposals are evaluated through a transparent and open process. The proposals selected are driven forward in collaboration with scientists at Almirall that result in innovative actionable projects.

In summary, accessing disruptive technologies and developing new treatment paradigms is key to the sustainability of R&D. Harnessing innovation at different stages of development with different partnership models will provide better outcomes for patients.

REFERENCES

- Schweizer, L. & He, J. Guiding principles of value creation through collaborative innovation in pharmaceutical research. *Drug Discov. Today* 23, 213–218 (2018).
- Holmes, D. A new chapter in innovation. *Nature Outlook: Open Innovation* 533, S54–S55 (2016).



almirallshare

AlmirallShare is an open innovation platform created as an opportunity to learn from each other and share knowledge, efforts, motivations and resources.

Are you a Scientist with an interest in dermatological research? Discover **AlmirallShare collaboration opportunities**

Adding value to your assets

Finding new uses in dermatology for existing advanced small molecules

Novel targets for skin health

Looking for novel targets and concepts for the treatment of dermatological diseases

Join us at sharedinnovation.almirall.com